

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 9.50981 Seconds
(without alignments)
2953.573 Million cell updates/sec

Title: US-10-650-585-15
Perfect score: 1532
Sequence: 1 AAGLIRACMLVRKAGHYV.....RGVAKAVDFIPVSEMTWR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1490	97.3	3010	1 A45573	genome polyprotein
2	1473	96.1	3010	1 GNMVWJ	genome polyprotein
3	1463	95.5	3010	1 GNMVWJ	genome polyprotein
4	1434	93.6	3010	1 S18030	genome polyprotein
5	1424	93.0	3010	1 GNMVWC	genome polyprotein
6	1360	88.8	3011	1 S40770	genome polyprotein
7	1355	88.4	3011	1 GNMVW3	genome polyprotein
8	1342	87.6	3011	1 GNMVW3	genome polyprotein
9	1192	77.8	3014	1 JCS620	genome polyprotein
10	1130	73.8	3033	1 JQ1303	genome polyprotein
11	1110	72.5	3033	1 GNMVW8	genome polyprotein
12	397.5	25.9	3005	2 T08835	polyprotein - dour
13	341	22.3	2970	2 T08835	polyprotein - dour
14	101	6.6	600	2 B46642	DNA-directed DNA p
15	97.5	6.4	1085	2 T03511	cohn protein homol
16	95.5	6.2	470	2 JCS620	tetracycline 6-hyd
17	93.5	6.1	353	2 G87392	conserved hypochet
18	93	6.1	660	1 VHMW2	structural protein
19	92.5	6.0	706	2 S33761	transferrin precu
20	92.5	6.0	716	2 G83612	hypothetical prote
21	91	5.9	904	2 A84212	hypothetical prote
22	90.5	5.9	868	2 H81775	aconitate hydratase
23	90	5.9	2796	2 JCS620	faty-acid synthase
24	89.5	5.8	7463	2 T36248	CDA peptid synthase
25	88	5.7	659	1 B44212	structural protein
26	87	5.7	3414	1 GNMVW8	genome polyprotein
27	86.5	5.6	3412	1 GNMVW8	genome polyprotein
28	85.5	5.6	470	1 GNMVW8	exo-alpha-sialidas
29			348	2 H70549	probable pdhb prot

30	85	5.5	470	1 NMTV9	exo-alpha-sialidas
31	85	5.5	707	2 D84154	cadmium-transport
32	84.5	5.5	1057	2 T38171	hugl protein - hum
33	84	5.5	233	2 T35594	hypothetical prote
34	84	5.5	3069	2 H70656	faty-acid synthase
35	83.5	5.5	299	2 AH3447	cytochrome-c oxida
36	83.5	5.5	315	2 AG2361	hypothetical prote
37	83.5	5.5	538	2 S23405	D-alanyl-D-alanine
38	83.5	5.5	990	2 S67499	glutamate synthase
39	83	5.4	398	2 B71284	probable periplasm
40	82.5	5.4	453	1 NMTV3	exo-alpha-sialidas
41	82.5	5.4	868	2 C81200	aconitate hydratase
42	82.5	5.4	5627	2 C83339	hypothetical prote
43	82	5.4	404	2 F87393	aminotransferase,
44	82	5.4	612	2 T01346	glucan 1,4-alpha-g
45	81.5	5.3	278	2 S51747	light harvesting c

ALIGNMENTS

RESULT 1
A45573 genome polyprotein - hepatitis C virus (strain UT)
N/Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C/Accession: A45573
R/Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, A
Virus Res. 23, 39-53, 1992
A/Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: se
A/Reference number: A45573; UID:92295714; PMID:1316627
A/Accession: A45573
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-3010 <TAN>
A/Cross-references: GB:D11168; GB:D01171; NID:9221612; PIDN:BA01943.1; PID:9221613
A/Experimental source: HCV-UT
A/Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:106207)
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EMP>
F/192-389/Product: major envelope protein E #status predicted <MEB>
F/390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: hepatitisin #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (2-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match	97.3%	Score 1490;	DB 1;	Length 3010;
Best Local Similarity	96.2%	Pred. No. 9.5e-12;		
Matches 281;	Conservative 5;	Mismatches 6;	Indels 0;	Gaps 0;
QY	1	AAGLIRACMLVRKAGHYVQMAFMKLAALNTYYDHLTPLODAHAGLRDLAAVAPV	60	
DB	915	AAGLIRACMLVRKAGHYVQMAFMKLAALNTYYDHLTPLODAHAGLRDLAAVAPV	974	
QY	61	ITSDMEVKIITGADTTAAGDITISGLPVASRRGRILLGPADNFGCGKRLIAPITAYSQ	120	
DB	975	VSDMEVKIITGADTTAAGDITISGLPVASRRGRILLGPADNFGCGKRLIAPITAYSQ	1034	
QY	121	QRRGLGCIITSLTRDNNOVEGEVQVSTATQSLATCVNGVCTVPHGAGSKTLAPK	180	
DB	1035	QRRGLGCIITSLTRDNNOVEGEVQVSTATQSLATCVNGVCTVPHGAGSKTLAPK	1094	
QY	181	GPITQYTNVDDLVGQAPGARSMTPTCGSSDPLVYTRADVYPPRRRDSGSLIS	240	

DB 1095 GPTGMYTNVDDLVGWHAPGARGSLPTCTGSSDLYLTVRHADVIPVRRGDSRGSLLS 1154

QY 241 PRPVSYLKGSSGGPILCPSSGHAVGIFRAAVCTRGVAKADPIPVESMETTMR 292

DB 1155 PRPVSYLKGSSGGPILCPSSGHAVGIFRAAVCTRGVAKADPIPVESMETTMR 1206

RESULT 2

GNMWUJ
genome polypeptide - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C:Accession: A39253; PMID:91088550; PMID:2175993
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Okoshi, S.; Sugimura, T.; Shimotoh
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A:Reference number: A39253; MUID:91088550; PMID:2175993
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <KAT>
A:Cross-references: GB:P00208; NID:G221610; PIDN:BA14233.1; PID:G221611
R:Kato, N.; Okoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A:Reference number: P00085
A:Accession: P00086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <KAT>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this genome polypeptide
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptide; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <MEP>
F:112-389/Product: nonstructural protein NS1 #status predicted <NS1>
F:390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F:730-1006/Product: nonstructural protein NS3 #status predicted <NS3>
F:1007-1615/Product: nonstructural protein NS4 #status predicted <NS4>
F:1230-1237/Product: nonstructural protein NS5 #status predicted <NS5>
F:1312-1317/Product: nucleotide-binding motif A (P-loop)
F:1316-1317/Product: nucleotide-binding motif B
F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,2240,2

Query Match 96.1%; Score 1473; DB 1; Length 3010;
Best Local Similarity 94.5%; Pred. No. 2,9e-119;
Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAPEV 60

DB 915 AAGLIRACMLVRKKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAPEV 974

QY 61 IFSMEVKIITWGDYTAACGDIISGLPVASARGSEIILGPADNPEGQMRILAPITVYSQ 120

DB 975 IFSMEVKIITWGDYTAACGDIISGLPVASARGSEIILGPADNPEGQMRILAPITVYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQVEGVYVSTATGSEFLATCVNGVCMVTFHAGSKTLAAGPK 180

DB 1035 QTRGLGCIITSLTGRDNQVEGVYVSTATGSEFLATCVNGVCMVTFHAGSKTLAAGPK 1094

QY 181 GPTGMYTNVDDLVGWHAPGARGSLPTCTGSSDLYLTVRHADVIPVRRGDSRGSLLS 240

DB 1095 GPTGMYTNVDDLVGWHAPGARGSLPTCTGSSDLYLTVRHADVIPVRRGDSRGSLLS 1154

QY 241 PRPVSYLKGSSGGPILCPSSGHAVGIFRAAVCTRGVAKADPIPVESMETTMR 292

DB 1155 PRPVSYLKGSSGGPILCPSSGHAVGIFRAAVCTRGVAKADPIPVESMETTMR 1206

RESULT 3

GNMWUJ
genome polypeptide - hepatitis C virus (strain Taiwan)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Note: host Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C:Accession: A40244
R:Chen, P.T.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
Virology 188, 102-113, 1992
A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
A:Reference number: A40244; MUID:92230206; PMID:1314449
A:Accession: A40244
A:Molecule type: genomic RNA
A:Residues: 1-3010 <CHE>
A:Cross-references: GB:M64754
C:Superfamily: hepatitis C virus genome polypeptide
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F:1-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: major envelope protein M #status predicted <MEP>
F:192-389/Product: nonstructural protein NS1 #status predicted <NS1>
F:390-729/Product: nonstructural protein NS2 #status predicted <NS2>
F:730-1006/Product: nonstructural protein NS3 #status predicted <NS3>
F:1007-1615/Product: nonstructural protein NS4 #status predicted <NS4>
F:1230-1237/Product: nonstructural protein NS5 #status predicted <NS5>
F:1312-1317/Product: nucleotide-binding motif A (P-loop)
F:1316-1317/Product: nucleotide-binding motif B
F:1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,

Query Match 95.5%; Score 1463; DB 1; Length 3010;
Best Local Similarity 93.8%; Pred. No. 2.1e-118;
Matches 274; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAPEV 60

DB 915 AAGLIRACMLVRKKAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGRLDAVAPEV 974

QY 61 IFSMEVKIITWGDYTAACGDIISGLPVASARGSEIILGPADNPEGQMRILAPITVYSQ 120

DB 975 IFSMEVKIITWGDYTAACGDIISGLPVASARGSEIILGPADNPEGQMRILAPITVYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQVEGVYVSTATGSEFLATCVNGVCMVTFHAGSKTLAAGPK 180

DB 1035 QTRGLGCIITSLTGRDNQVEGVYVSTATGSEFLATCVNGVCMVTFHAGSKTLAAGPK 1094

QY 181 GPTGMYTNVDDLVGWHAPGARGSLPTCTGSSDLYLTVRHADVIPVRRGDSRGSLLS 240

DB 1095 GPTGMYTNVDDLVGWHAPGARGSLPTCTGSSDLYLTVRHADVIPVRRGDSRGSLLS 1154

QY 241 PRPVSYLKGSSGGPILCPSSGHAVGIFRAAVCTRGVAKADPIPVESMETTMR 292

DB 1155 PRPVSYLKGSSGGPILCPSSGHAVGIFRAAVCTRGVAKADPIPVESMETTMR 1206

RESULT 4

GNMWUJ
genome polypeptide - hepatitis C virus (isolate JKI)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 23-Mar-2001
C:Accession: S18030; S33570; K48332; S18029
R:Honda, M.; Kaneko, S.; Maeshima, U.; Kobayashi, K.; Murakami, S.
Submitted to the EMBL Data Library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patier
A:Reference number: S18028
A:Accession: S18030
A:Molecule type: genomic RNA

A/Residues: 1-3010 <HON>
A/Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479
A/Experimental source: isolate JX1 from an individual
R/Honda, M.; Kaneo, S.; Uncura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A/Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A/Reference number: A48332; MUID:91119270; PMID:8360322
A/Accession: S33570
A/Molecule type: genomic RNA
A/Residues: 1-547, 'T', 549-621, 'V', 623-624, 'S', 626-652, 'DL', 655-761, 'T', 763-782 <HOW>
A/Cross-references: EMBL:X61591
A/Note: this sequence is inconsistent with the nucleotide translation
A/Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
as Trp, and TTC for residue 771 as Ser
A/Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:P121748)
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EPM>
F/192-389/Product: major envelope protein E #status predicted <MEB>
F/330-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (AS

Query Match 93.6%; Score 1434; DB 1; Length 3010;
Best Local Similarity 92.8%; Pred. No. 7e-116;
Matches 271; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTIVYDHLTPLOMAHAGRLDAVAPEV 60
Db 915 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTIVYDHLTPLOMAHAGRLDAVAPEV 974

QY 61 IFSDMEVKIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 120
Db 975 VFSDEMKIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQVEGVQVSTATQSEFLATCVNGVCMVTFHAGSKTLAPK 180
Db 1035 QTRGLGCIITSLTGRDNQVEGVQVSTATQSEFLATCVNGVCMVTFHAGSKTLAPK 1094

QY 181 GPITQYTNVDQDLVGMQAPPARSMTPTCGSSDLVLTFRADVIPTRRRDSGSLLS 240
Db 1095 GPITQYTNVDQDLVGMQAPPARSMTPTCGSSDLVLTFRADVIPTRRRDSGSLLS 1154

QY 241 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 5
GNWVTC
genome polyprotein - hepatitis C virus
N/Containing: capsid protein C; envelope protein M; hepatitisin (BC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C/Accession: A38465
R/Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.;
J. Virol. 65, 1105-1113, 1991
A/Title: Structure and organization of the hepatitis C virus genome isolated from human
A/Reference number: A38465; MUID:91140698; PMID:1847440
A/Accession: A38465
A/Molecule type: genomic RNA
A/Residues: 1-3010 <TRK>
A/Cross-references: EMBL:M58335; NID:G329770; PIDN:AAA72945.1; PID:G329771
C/Superfamily: hepatitis C virus genome polyprotein

C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EPM>
F/192-389/Product: major envelope protein E #status predicted <MEB>
F/330-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)
F/1312-1317/Region: nucleotide-binding motif B
F/1316-1319/Region: DEXH motif
F/1616-1862/Product: nonstructural protein NS4 #status predicted <NS4>
F/1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F/2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F/196,209,234,250,305,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,224

Query Match 93.0%; Score 1424; DB 1; Length 3010;
Best Local Similarity 92.1%; Pred. No. 5.2e-115;
Matches 269; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTIVYDHLTPLOMAHAGRLDAVAPEV 60
Db 915 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTIVYDHLTPLOMAHAGRLDAVAPEV 974

QY 61 IFSDMEVKIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 120
Db 975 VFSDEMKIITWGADTAACGDIISGLPVARSRGREILGPADNFGQGRLLAPITAYSQ 1034

QY 121 QTRGLGCIITSLTGRDNQVEGVQVSTATQSEFLATCVNGVCMVTFHAGSKTLAPK 180
Db 1035 QTRGLGCIITSLTGRDNQVEGVQVSTATQSEFLATCVNGVCMVTFHAGSKTLAPK 1094

QY 181 GPITQYTNVDQDLVGMQAPPARSMTPTCGSSDLVLTFRADVIPTRRRDSGSLLS 240
Db 1095 GPITQYTNVDQDLVGMQAPPARSMTPTCGSSDLVLTFRADVIPTRRRDSGSLLS 1154

QY 241 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPGSHAVGIFRAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 6
S40770
genome polyprotein - hepatitis C virus
N/Containing: capsid protein C; envelope protein M; hepatitisin (BC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C/Species: hepatitis C virus
C/Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C/Accession: S40770; PC1285
R/Okamoto, H.
submitted to the EMBL Data Library, March 1992
A/Reference number: S40770
A/Accession: S40770
A/Molecule type: genomic RNA
A/Residues: 1-513 <OKA>
A/Cross-references: EMBL:D10749; NID:G221511; PIDN:BA00705.1; PID:G221512
A/Experimental source: isolate HC-U1
C/Superfamily: hepatitis C virus genome polyprotein
C/Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F/2-115/Product: capsid protein C #status predicted <CPC>
F/116-191/Product: envelope protein M #status predicted <EPM>
F/192-389/Product: major envelope protein E #status predicted <MEB>
F/330-729/Product: nonstructural protein NS1 #status predicted <NS1>
F/730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F/1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>
F/1230-1237/Region: nucleotide-binding motif A (P-loop)

F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 88.8%; Score 1360; DB 1; Length 3011;

Best Local Similarity 85.6%; Pred. No. 1,9e-109; Indels 0; Gaps 0;

Matches 249; Conservative 22; Mismatches 20;

QY 2 QGLIRACMLVRKAAAGHYVOMAFKMLAALGTYYVDHLTPLODMAHAGRLDAVAVEPI 61
 DB 916 QGLIRACMLVRKAAAGHYVOMAFKMLAALGTYYVDHLTPLODMAHAGRLDAVAVEPI 975
 QY 62 FSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFGQGMRLAPITAYSOQ 121
 DB 976 FSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFGQGMRLAPITAYSOQ 1035
 QY 122 TRGLIGCIITSLTGRDKNQVGEVQVSTANOSFLATCVGVCMTVFHAGSKTLAPK 181
 DB 1036 TRGLIGCIITSLTGRDKNQVGEVQVSTANOSFLATCVGVCMTVFHAGSKTLAPK 1095
 QY 182 PITOMYTNVDDDLVGMWAPPARGSMPTCTGSSDLYLVTRHADYIPVRRRDSRGLSP 241
 DB 1096 PITOMYTNVDDDLVGMWAPPARGSMPTCTGSSDLYLVTRHADYIPVRRRDSRGLSP 1155
 QY 242 RPVSYLKSGSGGPIILCPGSAVGIFFRAVCTRGVAKAVDPIVPSMETTR 292
 DB 1156 RPVSYLKSGSGGPIILCPGSAVGIFFRAVCTRGVAKAVDPIVPSMETTR 1206

RESULT 7

GNMW3
 genome polyprotein - hepatitis C virus (strain HCV-1)
 N:contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001
 C:Accession: A39166; PQ0403; PQ0404
 R:Chao, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Col
 Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
 A:Title: Genetic organization and diversity of the hepatitis C virus.
 A:Reference number: A39166; MUID:91172826; PMID:1848704
 A:Accession: A39166
 A:Molecule type: mRNA
 A:Residues: 1-3011 <CHO>
 A:Cross-references: GB:M67321; NID:G329873; PIDN:AAA45676.1; PID:G329874
 R:Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
 J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A:Reference number: PQ0393; MUID:92268871; PMID:1316939
 A:Accession: PQ0403
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CHA>
 A:Cross-references: DDBJ:D10128
 A:Experimental source: Isolates E-b16
 A:Accession: PQ0404
 A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: Isolates E-b17
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:115-191/Product: envelope protein M #status predicted <EPM>
 F:191-389/Product: major envelope protein E #status predicted <MEE>
 F:389-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:729-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepacivirin #status predicted <NS3>
 F:1616-1862/Product: nucleotide-binding motif A (P-loop)
 F:1863-2013/Region: nucleotide-binding motif B
 F:2014-3011/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:1196-209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,224
 Query Match 88.4%; Score 1355; DB 1; Length 3011;

Best Local Similarity 85.2%; Pred. No. 5.2e-109; Indels 0; Gaps 0;

Matches 248; Conservative 23; Mismatches 20;

QY 2 QGLIRACMLVRKAAAGHYVOMAFKMLAALGTYYVDHLTPLODMAHAGRLDAVAVEPI 61
 DB 916 QGLIRACMLVRKAAAGHYVOMAFKMLAALGTYYVDHLTPLODMAHAGRLDAVAVEPI 975
 QY 62 FSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFGQGMRLAPITAYSOQ 121
 DB 976 FSDMEVKIITWGADTAACGDIISGLPVASARGREILLGPADNFGQGMRLAPITAYSOQ 1035
 QY 122 TRGLIGCIITSLTGRDKNQVGEVQVSTANOSFLATCVGVCMTVFHAGSKTLAPK 181
 DB 1036 TRGLIGCIITSLTGRDKNQVGEVQVSTANOSFLATCVGVCMTVFHAGSKTLAPK 1095
 QY 182 PITOMYTNVDDDLVGMWAPPARGSMPTCTGSSDLYLVTRHADYIPVRRRDSRGLSP 241
 DB 1096 PITOMYTNVDDDLVGMWAPPARGSMPTCTGSSDLYLVTRHADYIPVRRRDSRGLSP 1155
 QY 242 RPVSYLKSGSGGPIILCPGSAVGIFFRAVCTRGVAKAVDPIVPSMETTR 292
 DB 1156 RPVSYLKSGSGGPIILCPGSAVGIFFRAVCTRGVAKAVDPIVPSMETTR 1206

RESULT 8

GNMW3
 genome polyprotein - hepatitis C virus (strain H)
 N:contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructu
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to GenBank, July 1992
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>
 A:Cross-references: GB:M67463; NID:G329737; PIDN:AAA45534.1; PID:G329738
 R:Inchausti, G.; Zebadee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compari
 A:Reference number: A41546; MUID:92052256; PMID:1658800
 A:Contents: annotation
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:115-191/Product: envelope protein M #status predicted <EPM>
 F:191-389/Product: major envelope protein E #status predicted <MEE>
 F:389-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:729-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepacivirin #status predicted <NS3>
 F:1616-1862/Product: nucleotide-binding motif A (P-loop)
 F:1863-2013/Region: nucleotide-binding motif B
 F:2014-3011/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4a>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4b>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,2346

Query Match 87.6%; Score 1342; DB 1; Length 3011;

Best Local Similarity 84.5%; Pred. No. 7e-108; Indels 0; Gaps 0;

Matches 246; Conservative 25; Mismatches 20;

QY 2 QGLIRACMLVRKAAAGHYVOMAFKMLAALGTYYVDHLTPLODMAHAGRLDAVAVEPI 61

```

Db      916 QGLIRICALKARKINGHYVQMAIIKLGALITGTCVYNHLAPLBDPAHNGLDLAAVAEPV 975
Qy      62 FSDMEVKIITWGATTAACGDIISGLPVARSRGREILLGPADNFGCGWRLAPITAYSQ 121
Db      976 FSRRETKLITWGADTAACGDIISGLPVARSRGREILLGPADNFGCGWRLAPITAYSQ 1035
Qy      122 TRGLIGCIITSLTRDKNQVEGEVQVSTATOSFLATCVAGVCTVPHGAGSKTLAPFKG 181
Db      1036 TRGLIGCIITSLTRDKNQVEGEVQVSTATOTLACINGVCTVPHGAGSKTLAPFKG 1095
Qy      182 PITQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADYIPVRRGDSRGSLLSP 241
Db      1096 PVITQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADYIPVRRGDSRGSLLSP 1155
Qy      242 RPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 292
Db      1156 RPISYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 1206

```

RESULT 9

```

JC5620
genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Keywords: ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polypeptide; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:384-408/Region: hypervariable #status predicted
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: nonstructural protein NS2 #status predicted <NS3>
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DEXH motif
F:1617-1869/Product: nonstructural protein NS4 #status predicted <N4>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4b>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

```

```

Query March 77.8%; Score 1192; DB 1; Length 3014;
Best Local Similarity 72.6%; Pred. No. 7.8e-95;
Matches 212; Conservative 41; Mismatches 39; Indels 0; Gaps 0;

```

```

Qy      1 AGLIRACMLVRKAGSHYVQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAAVAEPV 60
Db      916 ARALLRLCLLAKHLVYKGYQAAHLHGRITGTYIYDHLAPMDMAASGRELVATEPI 975
Qy      61 IFSDEMEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFGCGWRLAPITAYSQ 120
Db      976 VFSMEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFGCGWRLAPITAYSQ 1035
Qy      121 QTRGLIGCIITSLTRDKNQVEGEVQVSTATOSFLATCVAGVCTVPHGAGSKTLAPFK 180
Db      1036 QTRGLIGCIITSLTRDKNQVEGEVQVSTATOTLACINGVCTVPHGAGSKTLAPFK 1095
Qy      181 GPITQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADYIPVRRGDSRGSLLS 240

```

```

Db      1096 GPVYQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADYIPARRRDSRGSLLS 1155
Qy      241 RPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 292
Db      1156 RPISYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 1207

```

RESULT 10

```

JC01303
genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Keywords: ATP; glycoprotein; hydrolyase; P-loop; polypeptide; serine proteinase; transme
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitisin #status predicted <NS3>
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:1967-2097/Region: interferon sensitivity determining #status predicted

```

```

Query March 73.8%; Score 1130; DB 1; Length 3033;
Best Local Similarity 69.2%; Pred. No. 1.9e-89;
Matches 202; Conservative 42; Mismatches 46; Indels 0; Gaps 0;

```

```

Qy      1 AGLIRACMLVRKAGSHYVQMAFMKLAALTGTYYVDHLPLODMAHAGRLDAAVAEPV 60
Db      919 AALLRMCTMVAHLHAGRVYQVALLARMTGTYYDHLTPMSDMAANGRLDAAVAEPI 978
Qy      61 IFSDEMEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFGCGWRLAPITAYSQ 120
Db      979 IFSPEMEVKIITWGADTAACGDIISGLPVARSRGREILLGPADNFGCGWRLAPITAYSQ 1038
Qy      121 QTRGLIGCIITSLTRDKNQVEGEVQVSTATOSFLATCVAGVCTVPHGAGSKTLAPFK 180
Db      1039 QTRGLIGCIITSLTRDKNQVEGEVQVSTATOTLACINGVCTVPHGAGSKTLAPFK 1098
Qy      181 GPITQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADYIPVRRRDSRGSLLS 240
Db      1099 GPITQMTYNTVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADYIPVRRRDSRGSLLS 1158
Qy      241 RPVSYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 292
Db      1159 RPISYLKSSGGPILCPGSHAVGIFRAAVCTRGVAKAVDPIVPSMETTMR 1210

```

RESULT 11

```

GNNV08
genome polyprotein - hepatitis C virus (strain HC-J8)
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
C:Species: hepatitis C virus
C:Keywords: ATP; glycoprotein; hydrolyase; P-loop; polypeptide; serine proteinase; transme
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitisin #status predicted <NS3>
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4 #status predicted <N4>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4b>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:1967-2097/Region: interferon sensitivity determining #status predicted

```

Virology 188, 331-341, 1992
 A>Title: Full-length sequence of a hepatitis C virus genome having poor homology to repo
 A/Reference number: A40250; MUID:9220232; PMID:1114459
 A/Accession: A40250
 A/Molecule type: genomic RNA
 A/Residues: 1-3033 <OK>
 A/Cross-references: GB:D10938; GB:D01221; NID:9221608; PIDN:BA01761.1; PID:9221609
 A/Chan, S.M.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L
 J. Gen. Virol. 73, 1131-1141, 1992
 A>Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A/Reference number: PQ0393; MUID:9226871; PMID:116939
 A/Accession: PQ0397
 A/Molecule type: genomic RNA
 A/Residues: 2678-2754 <CH>
 A/Cross-references: DDBJ:D10134
 A/Experimental source: isolate E-b12
 R/Kato, N.; Ootsuyama, Y.; Okoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnc
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991
 A>Title: Distribution of plural HCV types in Japan.
 A/Reference number: PQ0554; MUID:92068204; PMID:1720309
 A/Accession: PQ0559
 A/Molecule type: mRNA
 A/Residues: 2678-2728 <RAT>
 A/Cross-references: GB:D10562; GB:D05018; NID:9221523; PIDN:BA01418.1; PID:9221524
 A/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F.1.115/Product: capsid protein C #status predicted <CPC>
 F.1.16-191/Product: envelope protein M #status predicted <EMP>
 F.132-189/Product: major envelope protein E #status predicted <NS1>
 F.330-733/Product: nonstructural protein NS1 #status predicted <NS1>
 F.734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
 F.1011-1619/Product: nonstructural protein NS3 #status predicted <NS3>
 F.1334-1241/Region: nucleotide-binding motif A (P-loop)
 F.1334-1321/Region: nucleotide-binding motif B
 F.1320-1323/Region: DEH motif
 F.1320-1326/Product: nonstructural protein NS4 #status predicted <NS4>
 F.1867-2017/Product: nonstructural protein NS4b #status predicted <NS4>
 F.2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
 F.196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23
 Query Match 72.5%; Score 1111; DB 1; Length 3033;
 Best Local Similarity 67.8%; Pred. No. 8,7e-88;
 Matches 198; Conservative 44; Mismatches 50; Indels 0; Gaps 0;
 Db 1 AAGIIRACMLVRKAAGHYVOMAKALITGVYDHLTPLODMAHAGIRDAVAVERY 60
 919 AHAIRACTIVKHLAAGRYIOMLITIGRWTGYIYDHLSTWAAQGRDAIIVPEY 978
 Qy 61 IFSMEVXITTWGADTAACGDIISGLPVARSRGREITLGPADNFEQGMELAPITAYSQ 120
 Db 979 VFSMEKVVIVMGAEIVACGDIHLGLFVSARLGRVILGPADGTSKGMKLAPITAYQ 1038
 Qy 121 QTRGLICITLTLGRDKNOVEGVYVSTATOSFLATCVGVCTVFGAGSKTLAEPK 180
 Db 1039 QTRGLIAGIVSLTGRDKNOAGQVQLSSVTFQTLGTSISGLMTVYHAGNKTLAGPK 1098
 Qy 181 GPTQYMTNVDQDVGCAAPGARSMTCTCGSSDLYLTVRRHADVIVRRGSGRSGLS 240
 Db 1099 GPTQYMTNVAEGDVGWPSPTGKSLDPCGAGADLTVLRADVIVRRKDDRRGLLS 1158
 Qy 241 PRPVSYLKSSGSGPLTSPSGHVAIGIFRAAVCTRGVAAVDFIPVESMETTR 292
 Db 1159 PRPLSTLKSSGSGPVLCSRGHVAIGLFRRAVCAVAAASIDFIPVESLDAVTR 1210
 RESULT 12
 T08841
 polypotein - douroucouli hepatitis GB virus A
 C/Species: douroucouli hepatitis GB virus A
 C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
 C/Accession: T08841
 R/Erker, U.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998

A>Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A/Reference number: Z16486; MUID:99120818; PMID:9460920
 A/Accession: T08841
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-3005 <ERR>
 A/Cross-references: EMBL:AF023425; NID:92828599; PIDN:AA040502.1; PID:92828600
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: polyprotein

Query Match 25.9%; Score 397.5; DB 2; Length 3005;
 Best Local Similarity 34.8%; Pred. No. 9.4e-26;
 Matches 95; Conservative 46; Mismatches 119; Indels 13; Gaps 5;
 Db 23 AFMKLAALTYGVYDHLTPLODMAHAGIRDAVAVERYIPFSDEKVIITWGAADTAACGDI 82
 Db 887 AFVRLERSGVTLTHOQGQVSGAAAILXDLGVALLVEVSTAYARDCYIVRDAARTLACGR 946
 Qy 83 ISGLPVARSRRREILLG--PADNFEQGMELAPITAYSQOTRLLCIIITSLGRDKNQ 140
 Db 947 VEGLPVAVARRDDEVLVGFPSVRALPGCFVFAVVV--MORGLEFSSVVTSMIGRDERE 1005
 Qy 141 VEGVQVAVTAVQSFATCVGVCTVFGAGSKTLAEPKPIQMTNVDQDVGWCAQ 200
 Db 1006 HEGSLVUGTSTTSMGTCVGVCTVFGAGSKTLAEPKPIQMTNVDQDVGWCAQ 1065
 Qy 201 FGARSMTCTCGSSDLYLTVRRHADVIVRRGSGRSGLSFRPVSYLKSSGSGPLTSPG 260
 Db 1066 SGASGLPECKCGTOSVWCIRN--DGALCHGRLXVLEHDLPTETISDRGSSGSLTCDG 1123
 Qy 261 HAVGIFRAAVCTRGV-----AKAVDFIPVES 286
 Db 1124 HAVGM-VSLHKGKVTGVRVYKPEWETLPKDS 1155

RESULT 13
 T08839
 polypotein - marmoset hepatitis GB virus A
 C/Species: marmoset hepatitis GB virus A
 C/Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
 C/Accession: T08839
 R/Erker, U.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.
 J. Gen. Virol. 79, 41-45, 1998
 A>Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.
 A/Reference number: Z16486; MUID:99120818; PMID:9460920
 A/Accession: T08839
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: genomic RNA
 A/Residues: 1-2970 <ERR>
 A/Cross-references: EMBL:AF023424; NID:92828597; PIDN:AA040501.1; PID:92828598
 C/Superfamily: hepatitis C virus genome polyprotein
 C/Keywords: polyprotein

Query Match 22.3%; Score 341; DB 2; Length 2970;
 Best Local Similarity 30.5%; Pred. No. 7.6e-21;
 Matches 85; Conservative 46; Mismatches 96; Indels 52; Gaps 9;
 Db 45 WAHAG-----LRDLAVAVEPIFSDMEVKIITWGAADTAACGDIISGLPVARSRGRE 95
 Db 891 YAHAGQVTRRAEQLRQMGFALFEPVAVRBDCAVYRDAARTLSCGQVHGKPVVAAAGDE 950
 Qy 96 ILLGPADNFEQGMEL-----LAPITAYSQOTRLLCIIITSLGRDKNOVEGVYVS 149
 Db 951 VILGVLNV---WELPGFVPTAPVVVH--HHGKGFVGVVTSMTGMDDETHAVNVVVLG 1005
 Qy 150 TATQSFATCVGVCTVFGAGSKTLAEPKPIQMTNVDQDVGWCAQPGARSMTPC 209
 Db 1006 TSTRSMGTCVGVCTVFGAGSKTLAEPKPIQMTNVDQDVGWCAQPGARSMTPC 1065
 Qy 210 TCGSSDLYLTVRRHADVIVRRGSGRSGLS-----PPVSYLKSSGSGPLTSPG 258
 Db 1066 KQPGGVWVI-----RND--GALCHGRLXVLEHDLPTETISDRGSSGSLTCD 1112

QY 259 SGHVGITRAVCTRG-----YAKAVDFIPVESMETT 290
 ||||: :||| : :||: |
 Db 1113 EGHAVGML-ISYLHRSRVGTIRYTKPMETTPREALTHTT 1150

RESULT 14
PAGE 43

DNA-directed DNA polymerase (EC 2.7.7.7) alpha/DNA primase (EC 2.7.7.-) complex 68k chain
C/Species: Mus musculus (house mouse)
C/Date: 21-Sep-1993 #sequence revision 18-Nov-1994 #text change 20-Jun-2000
C/Accession: B46642
R/Miyazawa, H.; Irumi, M.; Tada, S.; Takada, R.; Masutani, M.; Ui, M.; Hanaoka, F.
J. Biol. Chem. 268, 8111-8122, 1993
A/Title: Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase α
A/Reference number: A46642; MUID:93216788; PMID:8463324
A/Accession: B46642
A/Status: Preliminary
A/Molecule type: mRNA, protein
A/Residues: 1-600 <MT>
A/Cross-references: GB:D13546; NID:G303658; PIDN:BA02746.1; PID:G303659
A/Experimental source: FMA3 cells
A/Note: sequence extracted from NCBI backbone (NCBIN:129148, NCBI:P:129149)
C/Keywords: nucleotidyltransferase

Query Match	6.6%;	Score 101;	DB 2;	Length 600;
Best Local Similarity	24.8%;	Pred. No. 0.77;		
Matches	55;	Conservative	34;	Mismatches 71;
				Indels 62;
				Gaps 12;

```

QY      L I S A C M L V R K A A G H V V Q V M - A F K I A A L L - - - - - G T Y V D H L - - - - - T P L D M A 46
  QY      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: 47
Db      27 L A E L C V L Y R O T E D G M S E L I A F C T S A G K C L L T V D I L N S F E F E V L N K L S K R A H S A S D S G 86
  QY      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: 87
Db      47 H A G I R D L A V A V E P V I F S D M E V K I I T W G A D I A A C G D I - I S G L P - - - - - V S A R G R E I 96
  QY      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: 97
Db      87 H A G R N D I - V S I G L I L A E E E E F L I S S Y T T P S K G P L K V S S T P E P L T K R S V A R S P R Q - 144
  QY      97 L I L G P A D N F E G C G K L L A P I T A N S O Q R G L L G C I I T S L G R D K N Q V E G V Q V S A T S F L 156
  QY      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: 157
Db      145 L I L S P S S - - - - - P S F S A T P S Q K - - - - - Y T S R T N R - - - - - G E V T T F G S A Q - - - 178
  QY      157 A T C N G V C W T V F H A S K T L - - A G P K G P I T Q M Y T W V D D L V G 196
  QY      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: 197
Db      179 - - - - - G L S M S G R G S G S V S L K V G D E P I T S G Y K A M F O O L M G 215

```

RESULT 15

10352
 10353
 10354
 10355
 10356
 10357
 10358
 10359
 10360
 10361
 10362
 10363
 10364
 10365
 10366
 10367
 10368
 10369
 10370
 10371
 10372
 10373
 10374
 10375
 10376
 10377
 10378
 10379
 10380
 10381
 10382
 10383
 10384
 10385
 10386
 10387
 10388
 10389
 10390
 10391
 10392
 10393
 10394
 10395
 10396
 10397
 10398
 10399
 10400
 10401
 10402
 10403
 10404
 10405
 10406
 10407
 10408
 10409
 10410
 10411
 10412
 10413
 10414
 10415
 10416
 10417
 10418
 10419
 10420
 10421
 10422
 10423
 10424
 10425
 10426
 10427
 10428
 10429
 10430
 10431
 10432
 10433
 10434
 10435
 10436
 10437
 10438
 10439
 10440
 10441
 10442
 10443
 10444
 10445
 10446
 10447
 10448
 10449
 10450
 10451
 10452
 10453
 10454
 10455
 10456
 10457
 10458
 10459
 10460
 10461
 10462
 10463
 10464
 10465
 10466
 10467
 10468
 10469
 10470
 10471
 10472
 10473
 10474
 10475
 10476
 10477
 10478
 10479
 10480
 10481
 10482
 10483
 10484
 10485
 10486
 10487
 10488
 10489
 10490
 10491
 10492
 10493
 10494
 10495
 10496
 10497
 10498
 10499
 10500
 10501
 10502
 10503
 10504
 10505
 10506
 10507
 10508
 10509
 10510
 10511
 10512
 10513
 10514
 10515
 10516
 10517
 10518
 10519
 10520
 10521
 10522
 10523
 10524
 10525
 10526
 10527
 10528
 10529
 10530
 10531
 10532
 10533
 10534
 10535
 10536
 10537
 10538
 10539
 10540
 10541
 10542
 10543
 10544
 10545
 10546
 10547
 10548
 10549
 10550
 10551
 10552
 10553
 10554
 10555
 10556
 10557
 10558
 10559
 10560
 10561
 10562
 10563
 10564
 10565
 10566
 10567
 10568
 10569
 10570
 10571
 10572
 10573
 10574
 10575
 10576
 10577
 10578
 10579
 10580
 10581
 10582
 10583
 10584
 10585
 10586
 10587
 10588
 10589
 10590
 10591
 10592
 10593
 10594
 10595
 10596
 10597
 10598
 10599
 10600
 10601
 10602
 10603
 10604
 10605
 10606
 10607
 10608
 10609
 10610
 10611
 10612
 10613
 10614
 10615
 10616
 10617
 10618
 10619
 10620
 10621
 10622
 10623
 10624
 10625
 10626
 10627
 10628
 10629
 10630
 10631
 10632
 10633
 10634
 10635
 10636
 10637
 10638
 10639
 10640
 10641
 10642
 10643
 10644
 10645
 10646
 10647
 10648
 10649
 10650
 10651
 10652
 10653
 10654
 10655
 10656
 10657
 10658
 10659
 10660
 10661
 10662
 10663
 10664
 10665
 10666
 10667
 10668
 10669
 10670
 10671
 10672
 10673
 10674
 10675
 10676
 10677
 10678
 10679
 10680
 10681
 10682
 10683
 10684
 10685
 10686
 10687
 10688
 10689
 10690
 10691
 10692
 10693
 10694
 10695
 10696
 10697
 10698
 10699
 10700
 10701
 10702
 10703
 10704
 10705
 10706
 10707
 10708
 10709
 10710
 10711
 10712
 10713
 10714
 10715
 10716
 10717
 10718
 10719
 10720
 10721
 10722
 10723
 10724
 10725
 10726
 10727
 10728
 10729
 10730
 10731
 10732
 10733
 10734
 10735
 10736
 10737
 10738
 10739
 10740
 10741
 10742
 10743
 10744
 10745
 10746
 10747
 10748
 10749
 10750
 10751
 10752
 10753
 10754
 10755
 10756
 10757
 10758
 10759
 10760
 1

Query Match	6.4%;	Score 97.5;	DB 2;	Length 1085;
Best Local Similarity	22.2%;	Pred. No. 3.2;		
Matches	82;	Conservative	30;	Mismatches 116;
				Indels 141;
				Gaps 18;

```

QY      14  AAGGHYQMAFMKIALTGYVYDHTP-----LDMAHAGLRDLAVAEVFLPSDMEV  67
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      144  AGGAVAAQALLAQLALASGLYA---PPIPGDKRLPDM---GFYPAAGVLPAREGSGPL  196
           :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      68  KIIW-----GADTACGDIIISGLEVSARRGREI---LLGPADNFBGQGMRLLAPI-  115

```

Db 197 AIVSFPRSYLTADTAPIDALIHGL-----BARGFACGLFAPSLKSEGGDALQJRLDG 251
 ::: ::: ::: ::: ::: :::
Qy 116 -----TANSOQ-----TRGL-----LGCI 129
 ::: ::: ::: ::: ::: :::
Db 252 RABALVWNATAFSAGADGTPFEDRYDVVFQVALSTARKEWMALADRGSLPADLAHV 311
 ::: ::: ::: ::: ::: :::
Qy 130 ITSJLGR-----DKQVEGEVGV-----STATQS 154
 ::: ::: ::: ::: ::: :::
Db 312 LPEVDGRLLPAGVVSFKAPRPDLEFRRFAHRDPRAVEALDRICGMHRLRQTPADR 371
 ::: ::: ::: ::: ::: :::
Qy 155 FLATCVN---GVCWTVFHCAGSKTLAGEKGPITQMTYNVDODLVGQAPPCARSMPCTC 211
 ::: ::: ::: ::: ::: :::
Db 372 KLAIVLSTPYGRSMQCAHVAHGDAIASTAEALLAQATE-----GYAIAFGALETITTA 425
 ::: ::: ::: ::: ::: :::
Qy 212 ----GSSDLIVLTRHADVIPARRRDSKSGSLSPRPVSYLKSSGGGLLCPSHAVGIFR 267
 ::: ::: ::: ::: ::: :::
Db 426 RLTWPLSD-VYLA-LDTLPEKLRAD-----LTPAAGPPBADPA-CADGAFH 468
 ::: ::: ::: ::: ::: :::
Qy 268 AAVCTRGVA 276
 ::: ::: ::: ::: ::: :::
Db 469 FAALRRGA 477

Search completed: May 6, 2004, 09:37:21
Job time : 10.5098.secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 6.09603 Seconds
(without alignments)
2494.160 Million cell updates/sec

Title: US-10-650-585-15

Perfect score: 1532
Sequence: 1 AAGLIRACMLVRRKAGSHV.....RGVAKAVDFIPVSMETTR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1490	97.3	3010	1	POLG_HCVUT
2	1473	96.1	3010	1	POLG_HCVUA
3	1463	95.5	3010	1	POLG_HCVTW
4	1424	93.0	3010	1	POLG_HCVXK
5	1355	88.4	3011	1	POLG_HCVI
6	1342	87.6	3011	1	POLG_HCVI
7	1130	73.8	3033	1	POLG_HCVU8
8	1111	72.5	3033	1	POLG_HCVU8
9	101	6.6	600	1	DPO2_MOUSE
10	93	6.1	660	1	DPO2_MOUSE
11	93	6.1	660	1	VST2_HEVVA
12	92.5	6.0	659	1	TRFE_HORSE
13	89.5	5.8	705	1	VST2_HEVME
14	88	5.7	3414	1	POLG_TBEVW
15	87	5.7	3412	1	POLG_TBEVW
16	87	5.7	3414	1	POLG_TBEVW
17	86.5	5.6	470	1	NRAM_IATRA
18	85	5.5	470	1	NRAM_IATRA
19	85	5.5	1705	1	HELS_MEIWA
20	84.5	5.5	1705	1	PTPV_MOUSE
21	84	5.5	309	1	UCP2_RAT
22	84	5.5	339	1	GPDA_CORF
23	84	5.5	470	1	NRAM_IATRA
24	83.5	5.5	538	1	DAC_ACTSP
25	83.5	5.5	538	1	DAC_ACTSP
26	82.5	5.4	453	1	NRAM_IATRA
27	82	5.4	309	1	UCP2_MOUSE
28	82	5.4	403	1	PCK_CHIMU
29	82	5.4	485	1	VST2_HEVVA
30	82	5.4	612	1	AMYG_ASPOR
31	81.5	5.3	660	1	VST2_HEVVA
32	81.5	5.3	398	1	TRMU_AGRIS
33	81.5	5.3	1022	1	CA26_CHICK

ALIGNMENTS

34	81	5.3	209	1	PAAD_PSEAE	Q9h08 pseudomonas
35	81	5.3	350	1	PE24_AARTH	Q9v04 arachidopsis
36	81	5.3	470	1	NRAM_IATRA	P03469 influenza a
37	81	5.3	730	1	HELS_MEIWA	Q8L13 methanocarc
38	80	5.2	309	1	UCP2_HUMAN	P55851 homo sapien
39	80	5.2	326	1	PANE_RHILLO	Q98705 rhizobium
40	80	5.2	3491	1	ERYL_SACER	Q03131 saccharopol
41	79.5	5.2	339	1	CRTB_RHOCA	P17056 rhodobacter
42	79.5	5.2	453	1	GAG_AYIMD	P06444 avian myelo
43	79.5	5.2	1123	1	PRYA_SOLTU	P30733 solanum tub
44	79	5.2	309	1	UCP2_PIG	O97562 sus scrofa
45	79	5.2	397	1	GALI_STRII	P13227 streptomyces

RESULT 1

POLG_HCVUT STANDARD; PRT; 3010 AA.

AC 000259;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polypeptide [contains: Capsid protein C (Core protein) (P22) ;

DE Envelope glycoprotein E1 (GP32) (GP35) ; Envelope glycoprotein E2

DE (GP68) (GP70) (NS1) ; Protein P7 ; Nonstructural protein NS2 (P21)

DE (EC 3.4.22.-) ; Protease/helicase NS3 (P70) (Hepatitisin)

DE (EC 3.4.21.98) ; Nonstructural protein NS4A (P4) ; Nonstructural protein

DE NS4B (P27) ; Nonstructural protein NS5A (P56) ; Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48) ;

OS Hepatitis C virus (isolate HC-UT) (HCV) .

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepatitisin.

OC NCBI_Taxid=11642;

OX NCBI_Taxid=11642;

RP SEQUENCE FROM N.A.

RX MEDLINE=92295714; PubMed=1318627;

RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.T.,

RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;

RT Molecular cloning of hepatitis C virus genome from a single Japanese

RT carrier: sequence variation within the same individual and among

RT infected individuals. ;

RL Virus Res. 23:39-53(1992).

-1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are

-1- hydrophobic, suggesting a possible membrane-related function. NS3

-1- and NS5 may play a role in the viral RNA replication.

-1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral

-1- precursor polypeptide, commonly with Asp or Glu in the P6

-1- position, Cys or Thr in P1 and Ser or Ala in P1'.

-1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +

-1- (RNA) (N).

-1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a

-1- lipoprotein envelope. The envelope consists of two proteins:

-1- protein M and glycoprotein E. The nucleocapsid is a complex of

-1- protein C and RNA.

-1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

-1- This SWISS-PROT entry is copyright. It is produced through a collaboration

-1- between the Swiss Institute of Bioinformatics and the EMBL outstation -

-1- the European Bioinformatics Institute. There are no restrictions on its

-1- use by non-profit institutions as long as its content is in no way

-1- modified and this statement is not removed. Usage by and for commercial

-1- entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

-1- or send an email to license@isb-sib.ch.)

CC EMBL; D11168; BAA01943.1; -

DR PIR; A45573; A45573.

DR MEROPS; S29.001; -

DR MEROPS; U39.001; -

DR InterPro; IPR003003; Cys_Ser_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV_capsid.

```
DR InterPro: IPR002521; HCV core.
DR InterPro: IPR002519; HCV env.
DR InterPro: IPR002531; HCV NS1.
DR InterPro: IPR002511; HCV NS2.
DR InterPro: IPR002745; HCV NS4a.
DR InterPro: IPR001490; HCV NS4b.
DR InterPro: IPR002868; HCV NS5a.
DR InterPro: IPR002166; HCV RdRp.
DR InterPro: IPR001650; Helicase C.
DR InterPro: IPR004109; Peptidase C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSV1r.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV env; 1.
DR Pfam: PF01560; HCV NS1; 1.
DR Pfam: PF01538; HCV NS2; 1.
DR Pfam: PF02907; HCV NS3; 1.
DR Pfam: PF01006; HCV NS4a; 1.
DR Pfam: PF01001; HCV NS4b; 1.
DR Pfam: PF01506; HCV NS5a; 1.
DR Pfam: PF00271; Helicase C; 1.
DR Pfam: PF00998; Viral RdRp; 1.
DR ProDom: PD16062; HCV NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KM Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KM Transmembrane; Nonstructural
INT _MET 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 326573 MW; 94A1C7435D642BB CRC64;
Query Match 97.3%; Score 1490; DB 1; Length 3010;
Best Local Similarity 96.2%; Pred. No. 6,7e-122;
Matches 281; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
```

```
DB 915 AGLIRACMLYKAVAGHYVQMAFMKLAALTGTGYVDHLTPLODMAHAGLDLAVAVEPV 974
QY 61 IFSDMEVETITMGADPTACGDIISGLPYASARGREIILGPADNDEGGWRLAPITAYSQ 120
DB 975 VFSDEMTETITMGADPTACGDIISGLPYASARGREIILGPADNDEGGWRLAPITAYSQ 1034
QY 121 QTRGLAGCIITSLTRGRDNQVEGEVQVSTATQSFATCVNGVCMYTFHGAASKTLAQP 180
DB 1035 QTRGLAGCIITSLTRGRDNQVEGEVQVSTATQSFATCVNGVCMYTFHGAASKTLAQP 1094
QY 181 GPITOMTNNVQDLVGMAPPGARSMPTCTGSSDYLVTNHAIVTPRRGDSRGLS 240
DB 1095 GPITOMTNNVQDLVGMAPPGARSMPTCTGSSDYLVTNHAIVTPRRGDSRGLS 1154
QY 241 PRPVSYLNGSSGGLPLCPGSHAVGIFRAVCTRGAVAVDFIPVESMETTR 292
DB 1155 PRPVSYLNGSSGGLPLCPGSHAVGIFRAVCTRGAVAVDFIPVESMETTR 1206
RESULT 2
ID POLG_HCVJA STANDARD; PRT; 3010 AA.
AC 26662;
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide (Coltains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepadiviruses.
OC NCBI_TaxID=11116;
RN [1]
RP MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).
RL [2]
RP MEDLINE=91088550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the Japanese hepatitis C viral genome.";
RT FEBS Lett. 280:325-328(1991).
RL [2]
CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
CC hydrophobic, suggesting a possible membrane-related function. NS3
CC and NS5 may play a role in the viral RNA replication.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
CC lipoprotein envelope. The envelope consists of two proteins:
CC protein M and glycoprotein E. The nucleocapsid is a complex of
CC protein C and mRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC EMBL: D90208; BAA14233.1; --
 DR PIR: A39253; GNMVCT.
 DR HSP: P26663; IUXP.
 DR MEROPS; S29.001; --
 DR MEROPS; U39.001; --
 DR InterPro: IPR009003; Cys Ser trypsin.
 DR InterPro: IPR00410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR00745; HCV_NS4.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RDRP.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; peptidase_C29.
 DR InterPro: IPR007095; RNA pol DS_PS.
 DR InterPro: IPR007094; RNA pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RDRP; 1.
 DR Pfam: PF0186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KM Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KM Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 KM INT1_MET 1
 FT CHAIN 1 115
 FT CHAIN 116 191
 FT CHAIN 192 383
 FT CHAIN 729
 FT CHAIN 1007
 FT CHAIN 1007 1615
 FT CHAIN 1616 1862
 FT CHAIN 1863 2013
 FT CHAIN 2014 3010
 FT TRANSMEM 347 369
 FT ACT SITE 1083 1083
 FT ACT SITE 1107 1107
 FT ACT SITE 1165 1165
 FT NP BIND 1230 1237
 FT SITE 1316 1319
 FT CAROXYD 196 196
 FT CAROXYD 209 209
 FT CAROXYD 234 234
 FT CAROXYD 250 250
 FT CAROXYD 305 305
 FT CAROXYD 417 417
 FT CAROXYD 423 423
 FT CAROXYD 440 440
 FT CAROXYD 448 448
 FT CAROXYD 532 532
 FT CAROXYD 556 556
 FT CAROXYD 576 576
 FT CAROXYD 623 623
 FT CAROXYD 645 645
 FT CAROXYD 2041 2041
 FT CAROXYD 2077 2077
 FT CAROXYD 2240 2240
 FT CAROXYD 2768 2768
 FT SEQUENCE 3010 AA; 327017 MM; AA939794F46DB185 CRC64;

Query Match 96.1%; Score 1473; DB 1; Length 3010;
 Best Local Similarity 94.5%; Pred. No. 2,1e-120;
 Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLRACMLVRAKAGHYVQMAFMKLAALGTGYVDHLTPLODAHAAGRLAVAVPV 60
 DB AAGLRACMLVRAKAGHYVQMAFMKLAALGTGYVDHLTPLODAHAAGRLAVAVPV 974
 QY 61 IPSDEMKIITWADTAACDIIISGLPVASARGREIILGPADNFGQCMRLIAPITAVSQ 120
 DB 975 VFSDEMTKLTIGADTAACDIIISGLPVASARGREIILGPADNFGQCMRLIAPITAVSQ 1034
 QY 121 QTRGLIGCIITSLTRDKNQVEGVQVYSTATQSLATCNQVCMTVHGAGSKTLAPK 180
 DB 1035 QTRGLIGCIITSLTRDKNQVEGVQVYSTATQSLATCNQVCMTVHGAGSKTLAPK 1094
 QY 181 GPTQMTYVDDQDLYGMOAPGARSGMTPTCGSSDLYIVTRHADVIPYRRRDSRGLIS 240
 DB 1095 GPTQMTYVDDQDLYGMOAPGARSGMTPTCGSSDLYIVTRHADVIPYRRRDSRGLIS 1154
 QY 241 PPVSYLKGSSGSGPLICPSGHAAGIFPAAVCTRGYAKAVDFIVESMETMR 292
 DB 1155 PPISYLGKSSGSGPLICPSGHAAGIFPAAVCTRGYAKAVDFIVESMETMR 1206

RESULT 3
 ID POLG_HCVTM STANDARD; PRT; 3010 AA.
 AC P29846;
 AD 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [containing: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis virus)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 DE Hepatitis C virus (isolate Taiwan) (HCV).
 OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2220206; PubMed=1314449;
 RA Chen P.U., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 "The Taiwanese hepatitis C virus genome: sequence determination and
 RT mapping the 5' terminus of viral genomic and antigenomic RNA.";
 RL Virology 188:102-113 (1992).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 hydrophobic, suggesting a possible membrane-related function. NS3
 and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 precursor polyprotein, commonly with Asp or Glu in the P6
 position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 (RNA) (N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and mRNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

RX MEDLINE=98227846; PubMed=9568891;
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,
 RA Steindler C., Tomei L., de Francesco R., Xie L.C., Chen Z.;
 RT "Complex of NS3 protease and NS4A peptide of HX strain hepatitis C
 RT virus: a 2.2 Å resolution structure in a hexagonal crystal form";
 RL Protein Sci. 7:837-847(1998).
 CC -1- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are
 CC hydrophobic, suggesting a possible membrane-related function. NS3
 CC and NS5 may play a role in the viral RNA replication.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: The virion of this virus is a nucleocapsid covered by a
 CC lipoprotein envelope. The envelope consists of two proteins:
 CC protein M and glycoprotein E. The nucleocapsid is a complex of
 CC protein C and RNA.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M58335; AA72945.1; -;
 DR PIR: A38465; GNMVTC.
 DR PDB: 1A10; 25-MAR-98.
 DR PDB: 1UXE; 14-JAN-98.
 DR PDB: 1NS3; 08-APR-98.
 DR PDB: 1C2P; 15-NOV-00.
 DR PDB: 1G5T; 08-NOV-99.
 DR PDB: 1GX5; 09-APR-02.
 DR PDB: 1QVU; 26-JUN-00.
 DR PDB: 1OHM; 20-APR-99.
 DR MEROPS: S29.001; -;
 DR MEROPS: U39.001; -;
 DR InterPro: IPR009003; Cys_ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR000745; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4A.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRP.
 DR InterPro: IPR004109; peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_env; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4A; 1.
 DR Pfam: PF01001; HCV_NS4B; 1.
 DR Pfam: PF01506; HCV_NS5A; 1.
 DR Pfam: PF00998; Viral_RdRP; 1.
 DR Pfam: PF01862; HCV_NSI; 1.
 DR SMART; SMO0482; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.

FT CHAIN	1	115	CAPSID PROTEIN C (POTENTIAL).
FT CHAIN	116	191	MATRIX PROTEIN (POTENTIAL).
FT CHAIN	192	333	MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN	384	729	NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN	730	1006	NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN	1007	1615	PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN	1616	1862	NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN	1863	2013	NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN	2014	3010	RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM	347	369	POTENTIAL.
FT ACT_SITE	1083	1083	CHARGE RELAY SYSTEM.
FT ACT_SITE	1107	1107	CHARGE RELAY SYSTEM.
FT ACT_SITE	1165	1165	CHARGE RELAY SYSTEM.
FT NE_BIND	1230	1237	ATP (POTENTIAL).
FT SITE	1316	1319	DECH BOX.
FT CARBOHYD	196	196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	209	209	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	305	305	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	417	417	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	423	423	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	430	430	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	448	448	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	540	540	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	623	623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2041	2041	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2077	2077	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2240	2240	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2529	2529	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	2788	2788	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1031	1035	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1039	1047	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1050	1050	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1059	1063	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1066	1074	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1075	1076	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1077	1081	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1082	1085	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1086	1087	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1090	1092	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1093	1094	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1095	1097	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1101	1103	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1104	1107	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1108	1112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1120	1120	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1122	1132	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1129	1133	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1135	1136	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1139	1144	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1149	1157	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1158	1161	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1162	1163	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1165	1166	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1168	1171	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1172	1174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1175	1186	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1187	1188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND	1189	1197	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT HELIX	1198	1202	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1203	1204	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN	1680	1688	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE	3010 AA;	327189 MM;	F8422D5ECCFPD9C CRC64;

Query Match 93.0%; Score 1424; DB 1; Length 3010;
 Best Local Similarity 92.1%; Pred. No. 4e-116;
 Matches 269; Conservative 10; Mismatches 13; Indels 0; Gaps 0;

Query Match 88.4%; Score 1355; DB 1; Length 3011;
 Best Local Similarity 85.2%; Pred. No. 4.4e-110;
 Matches 248; Conservative 23; Mismatches 20; Indels 0; Gaps 0;

QY 2 QGLIRACMLYKKAAGHYVQMAFMKALATGTYYDHLTFLQWMAAGLDLAVAEVPI 61
 DB 916 QGLIRFCALARKIMGHVYQVWVIRIKALGTGYVYNNHLPDRVANHGLDLVAVAEV 975
 QY 62 FSDMEVITITWGAADTAACGDISGLPVSARGREIILGPADNFEQCGMRLAPITAYSOQ 121
 DB 976 FSCQMETLITMGADTAACGDIINGLPVSARGREIILGPADNFEQCGMRLAPITAYSOQ 1035
 QY 122 TRGLIGCITITSLTRDKXQVEGEVQVSTATQSLATCVNGCVTFVHAGSKTLAPKG 181
 DB 1036 TRGLIGCITITSLTRDKXQVEGEVQVSTATQSLATCVNGCVTFVHAGSKTLAPKG 1095
 QY 182 PITGMVTVNODIYGMQAPGARGMTCTGSSDLVYTHADVTPRRRGDSRGLSP 241
 DB 1096 PVIQMTNVDDDLGMPAPQSSRLPCTCSSLVYTHADVTPRRRGDSRGLSP 1155
 QY 242 RPVSYLKSSGSGPLCPSGHNAVGIFFRAVCTRGVAKAVDFIPVESMETNR 292
 DB 1156 RPISYLKGSSGSGPLCPAGHAVGIFRAVCTRGVAKAVDFIPVENMETNR 1206

RESULT 6
 POLG HCVH STANDARD; PRT; 3011 AA.

AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Genome polypeptide (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP32); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.21.98); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 CC Hepacivirus.
 CX NCBI_Taxid=11106;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658600;
 RA Inchausti G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=9731332; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Mucke M.A., Lin C., Caron P.R.;
 RT Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN

CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
 CC (RNA) (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sib.ch).
 CC -----
 CC EMBL: M67463; AAA5534.1; -.
 CC PIR: A36814; GNVVCH
 CC PDB: 1HEI; 25-NOV-98.
 CC PDB: 1AIV; 16-FEB-99.
 CC PDB: 1AIR; 17-JUN-98.
 CC MEROPS: S29.001; -.
 CC MEROPS: U39.001; -.
 CC TRANSFAC: T04155; -.
 CC InterPro: IPR009003; Cys_Set_trypsin.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002542; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_enu.
 CC InterPro: IPR002531; HCV_NSI.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR000745; HCV_NS4A.
 CC InterPro: IPR001490; HCV_NS4B.
 CC InterPro: IPR002868; HCV_NS5A.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR001650; Helicase_C.
 CC InterPro: IPR004109; Peptidase_C29.
 CC InterPro: IPR007095; RNA_pol_ds_PS.
 CC InterPro: IPR007094; RNA_pol_PstI.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_enu; 1.
 CC Pfam: PF01560; HCV_NSI; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4A; 1.
 CC Pfam: PF01001; HCV_NS4B; 1.
 CC Pfam: PF01506; HCV_NS5A; 1.
 CC Pfam: PF00271; helicase_C; 1.
 CC Pfam: PF00398; Viral_RdRP; 1.
 CC ProDom: PD180662; HCV_NSI; 1.
 CC SMART: SM00487; DEXDc1; 1.
 CC PolyProtein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 CC 3D-structure.
 CC INIT_MET 1 1
 CC CHAIN 1 191
 CC CHAIN 192 383
 CC CHAIN 384 746
 CC CHAIN 747 809
 CC CHAIN 810 1026
 CC CHAIN 1027 1657
 CC CHAIN 1658 1711
 CC CHAIN 1712 1972
 CC CHAIN 1973 2420
 CC REMOVED FROM CAPSID PROTEIN C BY THE
 CC CELLULAR AMINOPEPTIDASE.
 CC CAPSID PROTEIN C.
 CC ENVELOPE GLYCOPROTEIN E1.
 CC ENVELOPE GLYCOPROTEIN E2.
 CC PROTEIN P7.
 CC NONSTRUCTURAL PROTEIN NS2.
 CC PROTEASE/HELICASE NS3.
 CC NONSTRUCTURAL PROTEIN NS4A.
 CC NONSTRUCTURAL PROTEIN NS4B.
 CC NONSTRUCTURAL PROTEIN NS5A.

-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC 1-1 SIMILARITY. THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

[illegible]

FT	CARBOHYD	558		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
FT	CARBOHYD	578		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
FT	CARBOHYD	627		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
FT	CARBOHYD	649		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
FT	CARBOHYD	1091		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
FT	CARBOHYD	2036		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
FT	CARBOHYD	2359		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
FT	CARBOHYD	2811		N-LINKED	(GLCNAC.	.)	(POTENTIAL).
QQ	SEQUENCE	3033	AA:	3301277	MM:	1A173E7E3381ED1DA	CRC64:	

Query Match	72.5%;	Score 1111;	DB 1;	Length 3033;
Best Local Similarity	67.8%;	Pred. No. 1e-88;		
Matches 198;	Conservative 44;	Mismatches 50;	Indels 0;	Gaps 0;

[illegible]

```

CC RESULT 9
CC DP02_MOUSE
CC ID_DP02_MOUSE STANDARD; PRT; 600 AA.
AC p33t11;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).
GN POLA3
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.
RX MEDLINE=93216788; PubMed=8463324;
RA Miyakawa H., Izumi M., Tada S., Takada R., Maenanti M., Ui M.,
RA Hanaka F.;
RT Molecular cloning of the cDNAs for the four subunits of mouse DNA
RT polymerase alpha-primease complex and their gene expression during
RT cell proliferation and the cell cycle.";
RL J Biol. Chem. 268:8111-8122(1993).
CC -I- FUNCTION: May play an essential role at the early stage of
CC chromosomal DNA replication by coupling the polymerase
CC alpha/primease complex to the cellular replication machinery (By
CC similarity).
CC -I- SUBUNIT: DNA polymerase alpha-primease is a four subunit enzyme
CC (subunits A, B, C and D), which is assembled throughout the cell
CC cycle. The largest subunit (subunit A) has DNA polymerase
CC activity, the two smallest subunits (subunits C and D) have DNA
CC primease activity. Subunit B binds to subunit A.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- PTM: PHOSPHORYLATED IN A CELL CYCLE DEPENDENT MANNER, IN G2/M
CC PHASE (BY SIMILARITY).
CC -I- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL/D13546; BA02746.1; -

DR PIR; B46642; B46642.

DR MGI; 99680; Polaz.

DR InterPro; IPR007200; DNA_pol_alpha_B.

DR Pfam; PF04058; DNA_pol_alpha_B; 1.

KM DNA replication; Nuclear protein; Phosphorylation.

FT DOMAIN 101 107 POLY-GLU.

FT PRO/SER/THR-RICH (HYDROPHILIC).

SEQUENCE 600 AA; 66267 MW; 79F94BEEF3FEBC CRC64;

Query Match 6.6%; Score 101; DB 1; Length 600;

Best Local Similarity 24.8%; Pred. No. 0.36;

Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 4 LIBRACMLVRKAGGHVYQW-AFMELALT-----GTYVDHL-----TPQDMA 46

DB 27 LAECVLYRQEDMGSELINFTSAGKTCITDIINSFEYVLNKLSTAMHSASKDSG 86

QY 47 HAGLRDLAVAPVIFSDMEVKITWGADTAACGDI--ISGLP-----VSARRGREI 96

DB 87 HAGLRDL-VSIQELIEAEEBEETLLSYTPSPKPLKRVSTPETPLTKSVARSPRQ- 144

QY 97 LIGGADNFEQGGWMLAPITVYSGQTGLGCITSLTGDKXQVEEYQVSTANQSL 156

DB 145 LLSPPS-----FSPSATPSQK-----YTSRTNR-----GEVVTTRGSAQ-- 178

QY 157 ATCVNGVCMVTFHAGSKTL--AGPKPITQMTYNTVQDVLG 196

DB 179 -----GLSMGSRGGSGSVSLKVVDPEPLTGSYKAMQQLMG 215

RESULT 10

VST2 HEVBU STANDARD; PRT; 660 AA.

AC P23326;

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)

DE Structural protein 2 precursor (ORF2).

OS Hepatitis E virus (strain Burma) (HEV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage;

OC Hepatitis E-like viruses.

NCBI_TaxID=31767;

OX NCBI_TaxID=31767;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=92024067; PubMed=1926770;

RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,

RY Fry K.E., Reyes G.R.;

RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the

full-length viral genome";

VL Virology 185:120-131(1991).

-1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING

THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA

BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M73318; AAA45736.1; -

DR PIR; C40778; VHWKH2.

DR InterPro; IPR004261; SP2.

DR InterPro; IPR008975; Viral_cap_coat.

DR Pfam; PF03014; SP2; 1.

KM Signal. 1 19 POTENTIAL.

FT SIGNAL. 20 660 STRUCTURAL PROTEIN 2.

FT CHAIN 188 660 5832A013CCCA461C CRC64;

SEQUENCE 660 AA; 70978 MW; 5832A013CCCA461C CRC64;

Query Match 6.1%; Score 93; DB 1; Length 660;

Best Local Similarity 19.0%; Pred. No. 2;

Matches 72; Conservative 47; Mismatches 119; Indels 140; Gaps 16;

QY 1 AGLLRACMLVRKAGGHVYQWAFMKALALTTGYVDHLTPLODPAHAGLRDLAVAPV 60

DB 188 AATATRYRDLVFNNAAGVAISISFWPQTTPTSV-----DNMSITSTVRLVDPG 239

QY 61 IFSDMEVKITWGADTAACGDIISGLPVASRRGRIILGPAD--NFEQGWMLAPI-TA 117

DB 240 IASELYI-----PSERLHYRQGRASVETSGVA 267

QY 118 YSQQTRGLD-----GCITSLTG----- 135

DB 268 EEBATSGILVWLCIHGSLVSYNTPTYGALGLDPALEPRNLTPGNTTRVSRYSPTA 327

QY 136 --RDKQVGEVQVSTAFSLA---TCVNGV-----CWTFH----- 169

DB 328 RHRLRGADGTALVTTTATRFMKDLYFTSTNGVEIGRIALTLPLNADTLGLLPTEL 387

QY 170 --GAG-----SKTLGPKG-PITQMTYNTVDQDVLGQAPGARSMTPCTCGSSDLYLV-- 219

DB 388 ISAGGGLFYSRPVYSANEFVTKLYTSVENA-----QQKGLAIPHDIDLGSRVLIQDY 443

QY 220 -TRHADVIVRRRGDSRG-SLSPRPVSYLK-----GSSGGLPLCSGARVGF 266

DB 444 DNOHEQDRPTPAPRPFSVLRANDVWLSTLAAYDOSTYSGSTGPYV--SDSVTLV 501

QY 267 RAAVCTRGVAKAVDFIPV 284

DB 502 NVATGQAVARSLDWTKV 519

RESULT 11

VST2 HEVPA STANDARD; PRT; 660 AA.

AC P33436;

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)

DE Structural protein 2 precursor (ORF2).

OS Hepatitis E virus (strain Pakistan) (HEV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage;

OC Hepatitis E-like viruses.

NCBI_TaxID=33774;

OX NCBI_TaxID=33774;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=92115700; PubMed=1731327;

RA Tareev S.A., Emerson S.U., Reyes G.R., Tareeva T.S., Legters L.J.,

RY Malik I.A., Iqbal M., Purcell R.H.;

RT "Characterization of a prototype strain of hepatitis E virus";

Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).

-1- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING

THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA

BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M80581; AAA45727.1; -

DR PIR; M80581; SP2.

DR InterPro; IPR004261; SP2.

DR InterPro; IPR008975; Viral_cap_coat.


```

QY 22 MAENKALALGTGVY--YDHLTPIDQMAHAGRLDVAVEPVFSMEVKITTWGA---- 74
DB 321 LGFRIIPAMDITWLYGYEYVT-----AIRNLREDIRPEVRKD-ECKKVKCAIGH 371
QY 75 DLAACGP--IISGLPVAGRRG-----ELLGPAANFEGOGRL-----LAPITAY 118
DB 372 EKVKCDWSTVNSGNIENESQSTEDCIATIKVGEADAMSIDGFFITIAKCGIYPLAE 431
QY 119 SQOTRGLGCIITSLTRDRKNQVGEVQVSTATQSEPLATCVNGVCTVPHGAGSKTLA 178
DB 432 NYEHRSGACVDTEEGYH-----AAVAVKSSDPDLT-----W-----NSLKG 470
QY 179 PKGITQMTYVDDDLVGMQAPPGARSMPTCGSSSLYLVTNRADYIPVRRGDSGSL 238
DB 471 KK---SCHTGVDR-TAGMNI PMGL-----LYSEIKICEFDKFFREGCAFGYR 513
QY 239 LSPRPVSLKSSGSGP-LIC-PSGHA-----VGIFRAVCTRGVAKAVDIPVESME 288
DB 514 RNSLTCLNLCIGSAGCPGRCPECPNHERYGYTGAFCLVEKGDVA-----FVKIQIYH 566

```

RESULT 13
VST2 HEVME STANDARD; PRT; 659 AA.

```

ID 003500;
AC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Structural protein 2 precursor.
OS Hepatitis E virus (strain Mexico) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxId=31768;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93079857; PubMed=1448913;
RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fy K.E.,
RA Bradley D.W., Tam A.W., Reyes G.R.;
RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
RT E virus (HEV).";
RL Virology 191:550-558(1992).
CC -!- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING
CC THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA
CC BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@sib.ch).
CC -----

```

```

DR EMBL: M74506; AAA45732.1; -
DR PIR: B44212; B44212.
DR InterPro: IPR004261; SP2.
DR InterPro: IPR008975; Viral_cap_coat.
DR Pfam: PF03014; SP2; 1.
KW Signal.
FT SIGNAL
FT CHAIN 23 659 STRUCTURAL PROTEIN 2.
SQ SEQUENCE 659 AA; 70640 MW; CF755755FDBEBC CRG4;

```

Query Match 5.8%; Score 89.5; DB 1; Length 659;
Best local similarity 17.8%; Pred. No. 4.1;
Matches 67; Conservative 50; Mismatches 121; Indels 139; Gaps 14;

```

QY 1 AAGIIRCMYKRAAGHYOMAKALATGYVYDHLTPIDQMAHAGRLDVAVEPV 60
DB 188 ARAIIRPPIVPNAVGVYAISIFWPQTTTPTSV-----DMSITSIDVRIIVQPG 239
QY 61 IFSDMEKIIITWGAADTAACGDIISGLPVASRRGEILLGPAD--NFGGQWRLL----- 112

```

```

DB 240 IASLVI-----PSERLHYRQGMESVETSGVA 267
QY 113 -----APIYVSGQ-----QTEGLGCIITSLTGR----- 136
DB 268 EEEATSGLYMCLHSGPVSNTPTVGTALGLDPALEEFNLTTCNNTVSAVSSA 327
QY 137 --DKQVGEVQVSTATQSEPLA---TCVNGV-----CWTVPHGAGSKTL 176
DB 328 RNSARGADGTAAITTTAATRFPMKDLHFTGLNCGVGVGIGIALTLNLADTLTGGLPTEH 387
QY 177 AGPKG-----PITQMTYVDDDLVGMQAPPGARSMPTCGSSSLYLVTNRADY 219
DB 388 SSAGGGLFYSRVPVANGEPYKLYTSVENA---QDQKVAIIPHDIDGDSRVIIQDYD 443
QY 220 TRHADYIPVRRGDSRG-SLSPRPVSYLK-----GSSGGLLCPGHAHVGIIPR 267
DB 444 NHEQDRPPTSPAPSPFSLRANDVLMSTLAEEVDQSTYGSSTQPYI--SDSVTLVN 501
QY 268 AAVCTRGVAKAVDIPYV 284
DB 502 VATGAQAVARSLDWSKV 518

```

RESULT 14
POLG TBEVM STANDARD; PRT; 3414 AA.

```

ID POLG_TBEVM
AC P14316; Q88493;
DT 01-JAN-1990 (Rel. 13, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Genome polyprotein [contains: Capsid protein C (Core protein); Matrix
DE protein (Envelope protein M); Major envelope protein E; Nonstructural
DE proteins NS1, NS2a, NS2b, NS4a and NS4b; Protease/helicase
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE (NS5)].
OS Tick-borne encephalitis virus (Western subtype) (TBEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxId=11088;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS.
RC STRAIN=Neudoerfl;
RX MEDLINE=96036491; PubMed=7483260;
RA Wallner G., Mandl C.W., Kunz C., Heinz F.X.;
RT "The flavivirus 3'-noncoding region: extensive size heterogeneity
RT independent of evolutionary relationships among strains of tick-borne
RT encephalitis virus.";
RL Virology 213:169-178(1995).
RN [2]
RP SEQUENCE OF 1-779 FROM N.A.
RC STRAIN=Neudoerfl;
RX MEDLINE=88322670; PubMed=3413985;
RA Mandl C.W., Heinz F.X., Kunz C.;
RT "Sequence of the structural proteins of tick-borne encephalitis virus
RT (Western subtype) and comparative analysis with other flaviviruses.";
RL Virology 166:197-205(1988).
RN [3]
RP SEQUENCE OF 767-3414 FROM N.A.
RC STRAIN=Neudoerfl;
RX MEDLINE=90051080; PubMed=2554575;
RA Mandl C.W., Heinz F.X., Stoeckl E., Kunz C.;
RT "Genome sequence of tick-borne encephalitis virus (Western subtype)
RT and comparative analysis of nonstructural proteins with other
RT flaviviruses.";
RL Virology 173:291-301(1989).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 281-660.
RX MEDLINE=95272700; PubMed=7753193;
RA Rey F.A., Heinz F.X., Mandl C.W., Kunz C., Harrison S.C.;
RT "The envelope glycoprotein from tick-borne encephalitis virus at 2-A
RT resolution.";
RL Nature 375:291-298(1995).
CC -!- FUNCTION: The small proteins NS2a, NS2b, NS4a and NS4b are

```



```

FT CHAIN 21129 1358 NONSTRUCTURAL PROTEIN NS2A (OR 1191).
FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS2B.
FT CHAIN 1490 2110 PROTEASE/HELICASE (NS3).
FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 2260 2510 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2511 3412 RNA-DIRECTED RNA POLYMERASE (NS5).
FT NP-BIND 1688 1695 ATP (POTENTIAL).
FT SITE 1779 1782 DEAD BOX.
FT TRANSMEM 101 112 HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
FT TRANSMEM 247 259 POTENTIAL.
FT TRANSMEM 266 280 POTENTIAL.
FT TRANSMEM 738 751 POTENTIAL.
FT DISULFID 283 310 BY SIMILARITY.
FT DISULFID 340 396 BY SIMILARITY.
FT DISULFID 354 385 BY SIMILARITY.
FT DISULFID 372 401 BY SIMILARITY.
FT DISULFID 466 570 BY SIMILARITY.
FT DISULFID 587 618 BY SIMILARITY.
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 983 983 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 999 999 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1228 1228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2447 2447 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2466 2466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 381 381 W -> S (IN REF. 3).
FT CONFLICT 850 850 E -> D (IN REF. 3).
SQ SEQUENCE 3412 AA; 377976. MW; 0F61C86DCCDC5965 CRC64;

```

Query Match 5.7%; Score 87; DB 1; Length 3412;

Best Local Similarity 24.4%; Pred. No. 51;

Matches 50; Conservative 23; Mismatches 58; Indels 74; Gaps 11;

```

QY 84 SGLVTSARRGRELLGPADNPEGQWRLAPITAYSOQTRGLGCIITSLTGRDKNQVEG 143
DB 1490 SGLVTSQGGREHEDRPEYKDV-YRIFSP-----GLM-----G 1524
QY 144 EVQY-VSTATQSFATCVNGCVTFPHAG--SKTLAGEKPTQMYTNVDOLV--- 195
DB 1525 QROGVGVGSKGVLAHT-----MHWVTRGAALSIDDAVAGP-----YMADVKEDVVCYGG 1573
QY 196 -----GMQA-----PRGARSMTPTCGSSDLVLTTRHADVIPIRRRGDSRGSILSP 241
DB 1574 AMSLEEKMKGETVOVHAHPFG-RAHEVHQCOPGELLDT-----GRRIGA 1617
QY 242 RPVSYLKSSGGPILCPGSHAVGIF 266
DB 1618 VPIDLAKGTSGSPILNSGVVGVLY 1642

```

Search completed: May 6, 2004, 09:31:55

Job time : 7.09603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 27.3102 Seconds

(without alignments)
3373.509 Million cell updates/sec

Title: US-10-650-585-15
Perfect score: 1532
Sequence: 1 AAGGIRACMVRKAGGHV.....RGVAKAVDFIPVSMETWR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1499	97.8	3010	12 Q9J3H7	Q9J3H7 hepatitis c
2	1499	97.8	3010	12 Q9D7D6	Q9D7D6 hepatitis c
3	1494	97.5	3010	12 P88803	P88803 hepatitis c
4	1494	97.5	3010	12 Q68826	Q68826 hepatitis c
5	1490	97.3	3010	12 Q9D7E6	Q9D7E6 hepatitis c
6	1490	97.3	3010	12 P90191	P90191 hepatitis c
7	1490	97.3	3013	12 Q9J3H4	Q9J3H4 hepatitis c
8	1489	97.2	3008	12 Q9J3F4	Q9J3F4 hepatitis c
9	1489	97.2	3010	12 Q9Q1Y3	Q9Q1Y3 hepatitis c
10	1486	96.9	3010	12 Q9J3F5	Q9J3F5 hepatitis c
11	1485	96.9	3010	12 Q9J3H5	Q9J3H5 hepatitis c
12	1485	96.9	3010	12 Q9D7E4	Q9D7E4 hepatitis c
13	1485	96.9	3010	12 Q807P3	Q807P3 hepatitis c
14	1482	96.7	3010	12 Q9J3I1	Q9J3I1 hepatitis c
15	1481	96.7	361	12 Q70822	Q70822 hepatitis c
16	1481	96.7	3010	12 Q9D7D7	Q9D7D7 hepatitis c

17	1481	96.7	3010	12 Q81541	Q81541 hepatitis c
18	1481	96.7	3011	12 Q9D7E3	Q9D7E3 hepatitis c
19	1480	96.6	361	12 Q70815	Q70815 hepatitis c
20	1480	96.6	3010	12 Q9J3G3	Q9J3G3 hepatitis c
21	1480	96.6	3010	12 Q9Q6P1	Q9Q6P1 hepatitis c
22	1480	96.6	3010	12 Q9J3H3	Q9J3H3 hepatitis c
23	1480	96.6	3010	12 Q9Q1Y4	Q9Q1Y4 hepatitis c
24	1480	96.6	3014	12 Q9D7E0	Q9D7E0 hepatitis c
25	1479	96.5	3010	12 Q9Q1Y7	Q9Q1Y7 hepatitis c
26	1479	96.5	3010	12 Q9J3I0	Q9J3I0 hepatitis c
27	1479	96.5	3010	12 Q9Q1Y8	Q9Q1Y8 hepatitis c
28	1479	96.5	3010	12 Q9J3H6	Q9J3H6 hepatitis c
29	1478	96.5	3010	12 Q9J3G6	Q9J3G6 hepatitis c
30	1478	96.5	3010	12 Q9Q1X5	Q9Q1X5 hepatitis c
31	1477	96.4	1186	12 Q81755	Q81755 hepatitis c
32	1477	96.4	2284	12 Q81817	Q81817 hepatitis c
33	1477	96.4	3010	12 Q9Q1X4	Q9Q1X4 hepatitis c
34	1477	96.4	3010	12 P89966	P89966 hepatitis c
35	1477	96.4	3010	12 Q9Q1X6	Q9Q1X6 hepatitis c
36	1477	96.4	3010	12 Q9Q1X5	Q9Q1X5 hepatitis c
37	1476	96.3	3010	12 Q9J3H9	Q9J3H9 hepatitis c
38	1476	96.3	3010	12 Q9J3H2	Q9J3H2 hepatitis c
39	1476	96.3	3010	12 Q9D7E0	Q9D7E0 hepatitis c
40	1476	96.3	3010	12 Q9Q1Y5	Q9Q1Y5 hepatitis c
41	1475	96.3	3010	12 Q9Q1Y5	Q9Q1Y5 hepatitis c
42	1475	96.3	3010	12 Q9D7E3	Q9D7E3 hepatitis c
43	1474	96.2	3010	12 Q68788	Q68788 hepatitis c
44	1473	96.1	3010	12 Q9Q1X2	Q9Q1X2 hepatitis c
45	1473	96.1	3010	12 Q9Q1X2	Q9Q1X2 hepatitis c

ALIGNMENTS

RESULT 1
ID Q9J3H7 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD15.
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease progression."
RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC
CC EMBL; AF207756; AAF65946.1; -
DR PIR; A61196; A61196.
DR PIR; PQ0246; PQ0246.
DR PIR; PQ0804; PQ0804.
DR PIR; PS0329; PS0329.
DR HSSP; P26663; 1UXP.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.

```

DR GO:0016740; F:transferase activity; IEA.
DR GO:0006118; P:electron transport; IEA.
DR GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO:0006350; P:transcription; IEA.
DR GO:0019079; P:viral genome replication; IEA.
DR GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys Ser_tyrp_sln.
DR InterPro: IPR00345; CysC_heme_BS.
DR InterPro: IPR01410; DEAD.
DR InterPro: IPR00522; HCV_capsid.
DR InterPro: IPR00521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5a.
DR InterPro: IPR004109; peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVlr.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR Pfam: PF018662; HCV_NS1; 1.
DR Pfam: PF018662; HCV_NS1; 1.
DR SMART; SMO0487; DEXdc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR Coated protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.
SQ SEQUENCE 3010 AA; 327365 MW; D86537317FFA106 CRC64;

Query Match 97.8%; Score 1499; DB 12; Length 3010;
Best Local Similarity 96.2%; Pred. No. 1.3e-124;
Matches 281; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGGIRACMVRKAGAGHYVQMAFMKLAALGTIVYDHLPLDMMARAGRLDAVAPEV 60
DB 915 AAGGIRACMVRKAGAGHYVQMAFMKLAALGTIVYDHLPLDMMARAGRLDAVAPEV 974
QY 61 IFSMEXVKITMGADTACGDIISGLFVSARGRHEILGADNFEQGMFLAPITAYSQ 120
DB 975 VFSDMETKITMGADTACGDIISGLFVSARGRHEILGADNFEQGMFLAPITAYSQ 1034
QY 121 QTRGLLCITSLGRKNGVEGVVSTATQSFATCVNGVCMVFHAGSXTLAGPK 180
DB 1035 QTRGLLCITSLGRKNGVEGVVSTATQSFATCVNGVCMVFHAGSXTLAGPK 1094
QY 181 GPITOMYTNVDQIVGQAPPGARSMPTCTGSSDLYLVTRHADVIPRRRGDSRGLLS 240
DB 1095 GPITOMYTNVDQIVGQAPPGARSMPTCTGSSDLYLVTRHADVIPRRRGDSRGLLS 1154
QY 241 PRPVSYLKSGSGGFLCPSGAVGIFRAAICTGVAACVAVFVESHMETMR 292
DB 1155 PRPVSYLKSGSGGFLCPSGAVGIFRAAICTGVAACVAVFVESHMETMR 1206

RESULT 2
Q9DPTD6 PRELIMINARY; FRT; 3010 AA.
AC Q9DPTD6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DS Genome polyprotein.
DS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

```

```

OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV221;
RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijioka M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited.";
RT Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; AB049101; BAB18514.1; -.
DR PIR; A61196; A61196.
DR PIR; P00246; P00246.
DR PIR; PS0329; PS0329.
DR HSSD; P26663; 1LXP.
DR GO:0016021; C:integral to membrane; IEA.
DR GO:0019028; C:viral capsid; IEA.
DR GO:0019031; C:viral envelope; IEA.
DR GO:0005524; F:ATP binding; IEA.
DR GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO:0005489; F:electron transporter activity; IEA.
DR GO:0016787; F:hydrolase activity; IEA.
DR GO:0003723; F:RNA binding; IEA.
DR GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO:0008236; F:serine-type peptidase activity; IEA.
DR GO:0005198; F:structural molecule activity; IEA.
DR GO:0016740; F:transferase activity; IEA.
DR GO:0006118; F:electron transport; IEA.
DR GO:0006508; F:proteolysis and peptidolysis; IEA.
DR GO:0006350; P:transcription; IEA.
DR GO:0019079; P:viral genome replication; IEA.
DR GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys Ser_tyrp_sln.
DR InterPro: IPR00345; CysC_heme_BS.
DR InterPro: IPR01410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRP.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVlr.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_RdRP; 1.
DR Pfam: PF018662; HCV_NS1; 1.
DR Pfam: PF018662; HCV_NS1; 1.
DR SMART; SMO0487; DEXdc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein; Transmembrane.
KW RNA-directed RNA polymerase; Transferrase; Transmembrane.
SQ SEQUENCE 3010 AA; 327108 MW; DE182D810E778BE4 CRC64;

Query Match 97.8%; Score 1499; DB 12; Length 3010;

```

Best Local Similarity 96.6%; Pred. No. 1,3e-124;
Matches 282; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLIRACMLVRKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
Db 915 AAGLIRACMLVRKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 61 IFSMEVKTITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAPITAYSQ 120
Db 975 VFSDETKITITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGCMTVPHGASKTLTAGPK 180
Db 1035 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGCMTVPHGASKTLTAGPK 1094
QY 181 GPITQMTNVDDIYGMQAPPGASMTPTCGSSDLVLTTRADVIIVRRRDSGSLLS 240
Db 1095 GPITQMTNVDDIYGMQAPPGASMTPTCGSSDLVLTTRADVIIVRRRDSGSLLS 1154
QY 241 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 3

P88803 PRELIMINARY; PRT: 3010 AA.

AC P88803;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1b;
RA Enomoto N.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-1b;
RA MEDLINE=95340824; PubMed=7542279;
RA Enomoto N., Sakuma I., Asahina Y., Kurosaki M., Murakami T.,
RA Yamamoto C., Izumi N., Marumo F., Sato C.;
RT "Comparison of full-length sequences of interferon-sensitive and
RT resistant hepatitis C virus 1b";
RL J. Clin. Invest. 96:224-230 (1995).
CC -1 SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL: D50484; BAA09075.1; -
DR PTR: A61196; A61196.
DR HSBP: P28663; INS3.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR GO: GO:0019028; C:Viral capsid; IEA.
DR GO: GO:0019031; C:Viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_ser_lypsin.
DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002522; HCV capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001430; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; peptidase_C29.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_Pol_r.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01586; HCV_NS5a; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00998; Viral_Rdrp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXdc1; 1.
KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1
FT CHAIN 192 383 E1.
FT CHAIN 384 809 E2.
FT CHAIN 810 1026 NS2.
FT CHAIN 1027 1657 NS3.
FT CHAIN 1658 1711 NS4a.
FT CHAIN 1712 1972 NS4b.
FT CHAIN 1973 2419 NS5a.
FT CHAIN 2420 3010 NS5b.
SQ SEQUENCE 3010 AA; 327332 MW; 5F81505783EFPB8 CRC64;

Query Match 97.5%; Score 1494; DB 12; Length 3010;

Best Local Similarity 96.2%; Pred. No. 3.6e-124;

Matches 281; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLIRACMLVRKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 60
Db 915 AAGLIRACMLVRKAAGHYVQMAFMKLAALGTYYVDHLTPLODMAHAGRLDAVAVEPV 974
QY 61 IFSMEVKTITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAPITAYSQ 120
Db 975 VFSDETKITITWGADTAACGDIISGLPVASRGREIILGPADNFEQGWRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGCMTVPHGASKTLTAGPK 180
Db 1035 QTRGLGCIITSLTGRDNQVGEVQVSTATOSFLATCVNGCMTVPHGASKTLTAGPK 1094
QY 181 GPITQMTNVDDIYGMQAPPGASMTPTCGSSDLVLTTRADVIIVRRRDSGSLLS 240
Db 1095 GPITQMTNVDDIYGMQAPPGASMTPTCGSSDLVLTTRADVIIVRRRDSGSLLS 1154
QY 241 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 292
Db 1155 PRPVSYLKSSGGPILCPSGHVGIFRAAVCTRGVAKAVDFIPVSMETMR 1206

RESULT 4

O68826 PRELIMINARY; PRT: 3010 AA.

AC O68826;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J33;
 RA Cho M.J.;
 RT "Molecular cloning of Hepatitis C virus genome from a single Japanese patient."
 RL Submitted (SEP-1991) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; D14484; BAA03375.1; -
 DR PIR; A61196; A61196.
 DR PIR; P00246; P00246.
 DR PIR; P00804; P00804.
 DR PIR; P50329; P50329.
 DR HSSP; P26663; IUXP.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0019031; C:Viral envelope; IEA.
 DR GO; GO:0008026; F:ATP binding; IEA.
 DR GO; GO:0005524; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0016787; F:Hydrolase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferrase activity; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; F:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral genome replication; IEA.
 DR InterPro; IPR009003; Cys_Ser_Trypsin.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_core.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_NS5a.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004109; Peptidase_C29.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NSI; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RdRP; 1.
 DR PRODOM; PD186062; HCV_NSI; 1.
 DR SMART; SM00487; DEXDC; 1.
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase; Hydroxylase; Nonstructural protein; Polypeptide; Transmembrane; RNA-directed RNA polymerase; Transferrase; Transmembrane.
 KW SEQUENCE 3010 AA; 327150 MW; 7270F4798455AFAD CRC64;

Query Match 97.5%; Score 1494; DB 12; Length 3010;
 Best Local Similarity 96.6%; Pred. No. 3.6e-124;
 Matches 282; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

1 AAGLIRACMLVRKAGHYVGMAMFKALALGTGVYDHLPLDQWALAGLRDLAVAEVP 60

DB 915 AAGLIRACMLVRKAGHYVGMAMFKALALGTGVYDHLPLDQWALAGLRDLAVAEVP 974
 QY 61 IFSDEVAIIITGADTAAGDIIISGLPVASARGREIILGPADNEGOGWELLAPITAYSQ 120
 DB 975 VFSDEMETIITWGADTAAGDIIISGLPVASARGREIILGPADNEGOGWELLAPITAYSQ 1034
 QY 121 QTRGLGCIITSLTRGRKNQVEGVQVSTQSFATCVNGVQWTFPHGSGKTLGPK 180
 DB 1035 QTRGLGCIITSLTRGRKNQVEGVQVSTQSFATCVNGVQWTFPHGSGKTLGPK 1094
 QY 181 GRITQMTNVNODLVGMQAPPGASMPCTCGSSDLVLTNRHADVIPVRERGDSRGLLS 240
 DB 1095 GRITQMTNVNODLVGMQAPPGASMPCTCGSSDLVLTNRHADVIPVRERGDSRGLLS 1154
 QY 241 PRPVSYLKGSGGPGILCPGSHAVGIFRAVCTRGVAKAVDPIPVESMETTR 292
 DB 1155 PRPVSYLKGSGGPGILCPGSHAVGIFRAVCTRGVAKAVDPIPVESMETTR 1206

RESULT 5
 Q9DTE6 PRELIMINARY; PRT; 3010 AA.
 AC Q9DTE6;
 DT 01-MAR-2001 (TRENDEL. 16, Created)
 DT 01-MAR-2001 (TRENDEL. 16, Last sequence update)
 DT 01-OCT-2003 (TRENDEL. 25, Last annotation update)
 DE Genome polypeptide.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV142;
 RA Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K., Hatanaka T., Ohta Y., Kanai K., Maruo H., Baba K., Hijioka M., Mishiro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 'progression score' revisited."
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL; AB049091; BAB18804.1; -
 DR PIR; A61196; A61196.
 DR PIR; P50329; P50329.
 DR HSSP; P26663; IUXP.
 DR GO; GO:0016021; C:Integral to membrane; IEA.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0019031; C:Viral envelope; IEA.
 DR GO; GO:0008026; F:ATP binding; IEA.
 DR GO; GO:0005524; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0008025; F:electron transporter activity; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferrase activity; IEA.
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006350; F:transcription; IEA.
 DR GO; GO:0019079; P:viral genome replication; IEA.
 DR GO; GO:0019087; P:viral genome replication; IEA.
 DR InterPro; IPR009003; Cys_Ser_Trypsin.
 DR InterPro; IPR000345; DEAD_heme_BS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NSI.

DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV_NS4A.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00480; HELICC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KM Hydroxylase; Nonstructural protein; Polypeptide; Transmembrane.
 KM RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327042 MW; 3807DC6879684C95 CRC64;

Query Match 97.3%; Score 1490; DB 12; Length 3010;
 Best Local Similarity 95.9%; Pred. No. 8; e-124;
 Matches 280; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
 QY 1 AAGLRACMVRKAGAGHYVOMAFMKLAALGTGVYDHLPLDPMAHAG;RDIAVAEPV 60
 Db 915 AAGLRACMVRKAGAGHYVOMAFMKLAALGTGVYDHLPLDPMAHAG;RDIAVAEPV 974
 QY 61 IFSMEVKIITWGDITACGDIISGLFVSARREHLLGPADNFEQGWELLAPITAYSQ 120
 Db 975 VFSMEVKIITWGDITACGDIISGLFVSARREHLLGPADNFEQGWELLAPITAYSQ 1034
 QY 121 QTRRLCCITSLTSLGRKNQVEGVYVATQSFATCNQGVWTFHAGSKTLAEPK 180
 Db 1035 QTRRLCCITSLTSLGRKNQVEGVYVATQSFATCNQGVWTFHAGSKTLAEPK 1094
 QY 181 GPITQWYTNVDQDLVGMQAPPGARSMTCTCGSSDLVLTNRHADVLPVRRRGRSGSLIS 240
 Db 1095 GPITQWYTNVDQDLVGMQAPPGARSMTCTCGSSDLVLTNRHADVLPVRRRGRSGSLIS 1154
 QY 241 PRPVSYLKSGSGGGLLCPGSHAVIFRAVCTRGVAQVDFIPVESHETMR 292
 Db 1155 PRPVSYLKSGSGGGLLCPGSHAVIFRAVCTRGVAQVDFIPVESHETMR 1206

RESULT 6

P90191 PRELIMINARY; PRT; 3010 AA.
 AC P90191;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HCV-1b;
 RA Enomoto N.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=HCV-1b;
 RX MEDLINE=95340824; PubMed=7542279;
 RA Enomoto N., Sakuma T., Asahina Y., Kurosaki M., Murakami T.,
 RA Yamamoto C., Izumi N., Maruno F., Sato C.;
 RT "Comparison of full-length sequences of interferon-sensitive and
 RT resistant hepatitis C virus 1b."
 RL J. Clin. Invest. 96:224-230(1995).
 CC -1- SUBMITTER: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 CC EMBL; D50482; BA09073.1; .
 DR PIR; A61196; A61196.
 DR PIR; P00254; P00254.
 DR PIR; P00804; P00804.
 DR PIR; P50329; P50329.
 DR PDB; 1DXW; 12-JAN-01.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0003668; F:RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006350; F:proteolysis and peptidolysis; IEA.
 DR GO; GO:0019079; F:viral genome replication; IEA.
 DR GO; GO:0019087; F:viral transformation; IEA.
 DR GO; GO:0009003; Cys_Ser_trypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KM Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 FT CHAIN 1 191
 FT CHAIN 192 383
 FT CHAIN 384 809
 FT CHAIN 810 1026
 FT CHAIN 1027 1657
 FT CHAIN 1658 1711
 FT CHAIN 1712 1972
 FT CHAIN 1973 2419
 FT CHAIN 2420 3010
 SQ SEQUENCE 3010 AA; 327438 MW; 5F15AC675A0C8268 CRC64;

Query Match

97.3%; Score 1490; DB 12; Length 3010;

Best Local Similarity 96.2%; Pred. No. 8,1e-124;
Matches 281; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRRKAGGHHYQMAFMKLAALTGTYYDHLTLPLOMAHAGRLDAVAEPV 60
Db 915 AAGLIRACMLVRRKAGGHHYQMAFMKLAALTGTYYDHLTLPLOMAHAGRLDAVAEPV 974
QY 61 IFSDMEVKIITWGDADTAACDIIISGLPVARSREIILGPADNFGQWRLLAFTAYSQ 120
Db 975 VFSDMEKIIITWGDADTAACDIIISGLPVARSREIILGPADNFGQWRLLAFTAYSQ 1034
QY 121 QTRGLGCIITSLTRGRDNQVEGEVYVSTATOSFLATCNGVCMTVHFGAGSTLAPK 180
Db 1035 QTRGLGCIITSLTRGRDNQVEGEVYVSTATOSFLATCNGVCMTVHFGAGSTLAPK 1094
QY 181 GPTOMTNNVODLVGMQAPPGARSMPCTCGSSDLYLVTRHADVI PVRRGRDSGLLS 240
Db 1095 GPTOMTNNVODLVGMQAPPGARSMPCTCGSSDLYLVTRHADVI PVRRGRDSGLLS 1154
QY 241 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
Db 1155 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 7
Q9J3H4 PRELIMINARY; PRT; 3013 AA.
ID Q9J3H4
AC Q9J3H4
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Genome polypeptide.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD18;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease progression."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
-1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA (BY SIMILARITY).
CC EMBL; AF207759; AAF65949.1; -
CC PIR; A61196; A61196.
DR PIR; P00804; P00804.
DR PIR; PS0329; PS0329.
DR HSP; P26663; TXP.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR003003; Cys Ser Lysln.
DR InterPro; IPR000345; Cys Ser Lysln.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV env.
DR InterPro; IPR002531; HCV NS1.
DR InterPro; IPR002518; HCV NS2.
DR InterPro; IPR000745; HCV NS4a.
DR InterPro; IPR001490; HCV NS4b.
DR InterPro; IPR002868; HCV NS5a.
DR InterPro; IPR002166; HCV NS5a.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004109; peptidase C29.
DR InterPro; IPR007095; RNA pol DS_PS.
DR InterPro; IPR007094; RNA pol_PSVlr.
DR Pfam; PF01543; HCV_capsid.1.
DR Pfam; PF01542; HCV_core.1.
DR Pfam; PF01539; HCV_env.1.
DR Pfam; PF01560; HCV NS1.1.
DR Pfam; PF01538; HCV NS2.1.
DR Pfam; PF02907; HCV NS3.1.
DR Pfam; PF01006; HCV NS4a.1.
DR Pfam; PF01001; HCV NS4b.1.
DR Pfam; PF01506; HCV NS5a.1.
DR Pfam; PF00271; Helicase_C.1.
DR Pfam; PF00998; Viral_RDRP.1.
DR ProDom; PD186062; HCV NS1.1.
DR SMART; SM00487; DEXDC.1.
DR PROSITE; PS00190; CYTOCHROME C.1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3013 AA; 326956 MW; E4F80FEA0E5C1E5 CRC64;

Query Match 97.3%; Score 1490; DB 12; Length 3013;
Best Local Similarity 96.2%; Pred. No. 8,1e-124;
Matches 281; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRRKAGGHHYQMAFMKLAALTGTYYDHLTLPLOMAHAGRLDAVAEPV 60
Db 915 AAGLIRACMLVRRKAGGHHYQMAFMKLAALTGTYYDHLTLPLOMAHAGRLDAVAEPV 974
QY 61 IFSDMEVKIITWGDADTAACDIIISGLPVARSREIILGPADNFGQWRLLAFTAYSQ 120
Db 975 VFSDMEKIIITWGDADTAACDIIISGLPVARSREIILGPADNFGQWRLLAFTAYSQ 1034
QY 121 QTRGLGCIITSLTRGRDNQVEGEVYVSTATOSFLATCNGVCMTVHFGAGSTLAPK 180
Db 1035 QTRGLGCIITSLTRGRDNQVEGEVYVSTATOSFLATCNGVCMTVHFGAGSTLAPK 1094
QY 181 GPTOMTNNVODLVGMQAPPGARSMPCTCGSSDLYLVTRHADVI PVRRGRDSGLLS 240
Db 1095 GPTOMTNNVODLVGMQAPPGARSMPCTCGSSDLYLVTRHADVI PVRRGRDSGLLS 1154
QY 241 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
Db 1155 PRPVSYLKGSSGGPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 1206

RESULT 8
Q9J3F4 PRELIMINARY; PRT; 3008 AA.
ID Q9J3F4
AC Q9J3F4
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Genome polypeptide.
GN MD34.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
CX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD34;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease progression."
CC EMBL; AF207759; AAF65949.1; -
CC PIR; A61196; A61196.
DR PIR; P00804; P00804.
DR PIR; PS0329; PS0329.
DR HSP; P26663; TXP.
DR GO; GO:0016021; C:Integral to membrane; IEA.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0008236; F:serine-type peptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR003003; Cys Ser Lysln.
DR InterPro; IPR000345; Cys Ser Lysln.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.

Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL: AF208024; AAE61205.1; -.
 DR PIR: A61196; A61196.
 DR PIR: PQ0246; PQ0246.
 DR PIR: PS0329; PS0329.
 DR HSP: P26663; IJXP.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0016787; F:hydrolase activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P:transcription; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_ser_lypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C29.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NSI; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; viral_RdRp; 1.
 DR Pfam: PF01806; HCV_NSI; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME C; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 KW Hydrolysis; Nonstructural protein; Polypeptide; Polypeptide;
 KW RNA-directed RNA polymerase; Transferase; Transmembrane.
 SO SEQUENCE 3008 AA; 326834 MW; 99AE09E14C3109F4 CRC64;

Query Match 97.2%; Score 1489; DB 12; Length 3008;
 Best Local Similarity 96.2%; Pred. No. 1e-123;
 Matches 281; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAGGYVGMFMKLAALGTGVVHLPICQMAHAGLDLAVAVEPV 60
 DB 913 AAGIIRACMLVRKAGGYVGMFMKLAALGTGVVHLPICQMAHAGLDLAVAVEPV 972

QY 61 IFSMNEVKIITWGDATACDIIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 120
 DB 973 VFSMNEVKIITWGDATACDIIISGLPVSARGREIILGPADNPEGQWMLAPITAYSQ 1032

QY 121 QTRGLGCIITSLTGRDNQVGEVQWSTATQSLATCVNGCVTFPHAGSKTLAPGR 180
 DB 1033 QTRGLGCIITSLTGRDNQVGEVQWSTATQSLATCVNGCVTFPHAGSKTLAPGR 1092

QY 181 GPITQYNTVDDDLVGMQAPPGARSMTPCTCGSSDLVLTIRADVIIPARRRDSRGSILS 240
 DB 1093 GPITQYNTVDDDLVGMQAPPGARSLTPCTCGSSDLVLTIRADVIIPARRRDSRGSILS 1152

QY 241 PRPVSYLKSGSGPLCPESGHAVGIFRAVCTRGVAKAVDFIPVSMETMR 292
 DB 1153 PRPVSYLKSGSGPLCPESGHAVGIFRAVCTRGVAKAVDFIPVSMETMR 1204

RESULT 9
 ID Q90IY3 PRELIMINARY; PRT; 3010 AA.
 AC Q90IY3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CC NCBI_TaxID=11103;
 CX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=MD4-2;
 RC MEDLINE=20013325; PubMed=10544098;
 RX Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
 RA Tazawa J.I., Izumi N., Marumo F., Sato C.;
 RT "Time-related changes in full-length hepatitis C virus and hepatitis
 RL virology 263:244-253(1999).
 RP [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MD4-2;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
 RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL: AF165052; AAD56187.1; -.
 DR PIR: A61196; A61196.
 DR PIR: PQ0246; PQ0246.
 DR PIR: PQ0804; PQ0804.
 DR PIR: PS0329; PS0329.
 DR HSP: P26663; IJXP.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P:transcription; IEA.
 DR GO: GO:0019079; P:viral genome replication; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_ser_lypsin.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NSI.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR000745; HCV_NS4a.

DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_NS5A.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C9.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327087 MW; 2FE78321686D4002 CRC64;

Query Match 97.2%; Score 1489; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 1e-123;
 Matches 279; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAEVP 60
 DB 915 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAEVP 974
 QY 61 IFSDEVKIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAPITAYSQ 120
 DB 975 VFSMEKVIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAPITAYSQ 1034
 QY 121 QTRGLGCIITSLTRGRNQVEGEVYSTATOSFLATCNVGVMTYFHGAGSKTLA 180
 DB 1035 QTRGLGCIITSLTRGRNQVEGEVYSTATOSFLATCNVGVMTYFHGAGSKTLA 1094
 QY 181 GPIQWNTYNDQDLVGMQAPGARSMPTCTCGSSDLYLVTRHADVIVRRRGDSRGLLS 240
 DB 1095 GPIQWNTYNDQDLVGMQAPGARSMPTCTCGSSDLYLVTRHADVIVRRRGDSRGLLS 1154
 QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFPAVCTRGVAKAVDFIPVESNETTMR 292
 DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFPAVCTRGVAKAVDFIPVESNETTMR 1206

RESULT 10

Q9J3F9 PRELIMINARY; PRT; 3010 AA.

AC Q9J3F9; OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_Taxid=1103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD33; Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
 RA Nagayama K., Characteristics of hepatitis C viral genome associated with disease
 RT "Characteristics of hepatitis C viral genome associated with disease
 RT progression";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL; AF207774; AAF65964.1; -.

DR PIR: A61196; A61196.
 DR PIR: P00246; P00246.
 DR PIR: P50329; P50329.
 DR HSSP: P27958; IHEI.
 DR MEROPS: S29.001; -.
 DR MEROPS: U39.001; -.
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0019028; C:viral capsid; IEA.
 DR GO: GO:0019031; C:viral envelope; IEA.
 DR GO: GO:0005524; F:ATP binding; IEA.
 DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F:electron transporter activity; IEA.
 DR GO: GO:0003723; F:RNA binding; IEA.
 DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F:structural molecule activity; IEA.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR GO: GO:0006118; F:electron transport; IEA.
 DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; F:transcription; IEA.
 DR GO: GO:0019079; F:viral genome replication; IEA.
 DR GO: GO:0019087; F:viral transformation; IEA.
 DR InterPro: IPR009003; Cys Ser-lysin.
 DR InterPro: IPR000345; Cys_Neme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002868; HCV NS5a.
 DR InterPro: IPR002166; HCV RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004109; Peptidase_C9.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327102 MW; 7162C9DB93B6E0C7 CRC64;

Query Match 97.0%; Score 1486; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 1.8e-123;
 Matches 279; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 1 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAEVP 60
 DB 915 AGLIRACMLVRKAGHYVQMAFMKLAALTGYVDHLTPLODMAHAGRLDAVAEVP 974
 QY 61 IFSDEVKIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAPITAYSQ 120
 DB 975 VFSMEKVIITWGADTAACGDIISGLPVSARREIILGPADNPEGQWLLAPITAYSQ 1034
 QY 121 QTRGLGCIITSLTRGRNQVEGEVYSTATOSFLATCNVGVMTYFHGAGSKTLA 180
 DB 1035 QTRGLGCIITSLTRGRNQVEGEVYSTATOSFLATCNVGVMTYFHGAGSKTLA 1094
 QY 181 GPIQWNTYNDQDLVGMQAPGARSMPTCTCGSSDLYLVTRHADVIVRRRGDSRGLLS 240


```

Db      1095 GPTQMTYNTVDLWGPAPPGARSLPCTCGSSDLVLRHNAVIVRRRRGSRGSLLS 1154
Qy      241 PRPVSYLKSSGGPFLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 292
Db      1155 PRPVSYLKSSGGPFLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 11
Q9J3H5 PRELIMINARY; PRT; 3010 AA.
ID Q9J3H5
AC Q9J3H5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
DE Hepatitis C virus.
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD17;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MENA (BY SIMILARITY).
DR EMBL: AF207758; AAF65948.1; -
DR PIR: A61196; A61196.
DR PIR: PQ0246; PQ0246.
DR PIR: PQ0254; PQ0254.
DR PIR: PS0329; PS0329.
DR HSSP: P27958; 1HE1.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019028; C:viral capsid; IEA.
DR GO: GO:0019031; C:viral envelope; IEA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO: GO:0005489; F:electron transporter activity; IEA.
DR GO: GO:0016787; F:hydrolase activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO: GO:0008286; F:serine-type peptidase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006118; F:electron transport; IEA.
DR GO: GO:0006508; F:proteolysis and peptidolysis; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR GO: GO:0019079; P:viral genome replication; IEA.
DR GO: GO:0019087; P:viral transformation; IEA.
DR InterPro: IPR009003; Cys_ser_lypsin.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NSI.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR004109; Peptidase_C39.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.

```

```

DR Pfam: PF01560; HCV_NSI; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NSI; 1.
DR SMART: SM00487; DEXDC; 1.
DR POSITE: PS00190; CYTOCHROME C; 1.
DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolyase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; transferase; Transmembrane.
SQ SEQUENCE 3010 AA; 326801 MW; 9FEE3D1B93B7AA4B CRC64;

Query Match          96.9%; Score 1485; DB 12; Length 3010;
Best Local Similarity 95.9%; Pred. No. 2,3e-123;
Matches 280; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy      1 AGLIPACMLVRKACGHTVQAFMKLALGTYYVDHITPLQDMAHAGLRDLAVAEV 60
Db      915 AGLIPACMLVRKAGHYQWALMKLALGTYYVDHITPLRDAHTGLRLAVAEV 974
Qy      61 IFSDMEVKIITWGADTAAAGDIIISGLPVASRGREILGPDNFEQGRILAPITAYGQ 120
Db      975 VESDMEKTIITWGADTAAAGDIIISGLPVASRGREILGPDNFEQGRILAPITAYGQ 1034
Qy      121 QTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFPLATCWGCVTFWAGSKTLAGPK 180
Db      1035 QTRGLIGCIITSLTGRDNQVEGEVQVSTATQSFPLATCINGCVTFWAGSKTLAGPK 1094
Qy      181 GPTQMTYNTVDLWGPAPPGARSLPCTCGSSDLVLRHNAVIVRRRRGSRGSLLS 240
Db      1095 GPTQMTYNTVDLWGPAPPGARSLPCTCGSSDLVLRHNAVIVRRRRGSRGSLLS 1154
Qy      241 PRPVSYLKSSGGPFLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 292
Db      1155 PRPVSYLKSSGGPFLCPGSHAVGIFRAVCTRGVAKAVDFIPVESMETMR 1206

RESULT 12
Q9JTE4 PRELIMINARY; PRT; 3010 AA.
ID Q9JTE4
AC Q9JTE4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV150;
RA Takahashi T., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Hatahara T., Ohta Y., Kanai K., Baba K., Hijikata M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MENA (BY SIMILARITY).
DR EMBL: AB049093; BAB18806.1; -
DR PIR: A61196; A61196.
DR PIR: PQ0246; PQ0246.
DR PIR: PQ0804; PQ0804.
DR PIR: PS0329; PS0329.
DR HSSP: P26663; 1UXP.
DR GO: GO:0016021; C:integral to membrane; IEA.

```

DR GO: GO:0019028; C: viral capsid; IEA.
 DR GO: GO:0019031; C: viral envelope; IEA.
 DR GO: GO:0005524; F: ATP binding; IEA.
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F: electron transporter activity; IEA.
 DR GO: GO:0003723; F: RNA binding; IEA.
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008286; F: serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F: structural molecule activity; IEA.
 DR GO: GO:0016740; F: transferase activity; IEA.
 DR GO: GO:0006118; F: electron transport; IEA.
 DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P: transcription; IEA.
 DR GO: GO:0019079; P: viral genome replication; IEA.
 DR GO: GO:0019087; P: viral transformation; IEA.
 DR InterPro: IPR003003; Cys_Ser_lysine.
 DR InterPro: IPR003451; CysC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002166; HCV NS5a.
 DR InterPro: IPR001650; HCV RdRp.
 DR InterPro: IPR004109; peptidase C29.
 DR InterPro: IPR007095; RNA_pol_DS_P8.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; viral RdRp; 1.
 DR Pfam: PF01662; HCV NS1; 1.
 DR Pfam: PF01662; HCV NS1; 1.
 DR SMART; SMO0487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR Coat protein, Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transmembrane.
 SQ SEQUENCE 3010 AA; 327324 MW; 3DE6CF249BD151C CRC64;

Query Match 96.9%; Score 1485; DB 12; Length 3010;
 Best Local Similarity 95.5%; Pred. No. 2.3e-123;
 Matches 279; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
 Db 1 AAGGIRACMVRKAGGAYVQMAFMKLAATGTYYVDHLPLODMARAGRLDAVAPEV 60
 915 AAGIRACMVRKAGGAYVQMAFMKLAATGTYYVDHLPLODMARAGRLDAVAPEV 974
 Db 61 IFSDMEVKITWAGDTAACGDIISGLFVSARGRREIILGPADNPEGQAFLLAPITAYSG 120
 975 VFSDMETKITWAGDTAACGDIISGLFVSARGRREIILGPADNPEGQAFLLAPITAYSG 1034
 Db 121 QTRGLGGLITSLGRKNGVEGVVSTATOSFLATCNGVCMWTFHAGAGSKTLAPGR 180
 1035 QTRGLGGLITSLGRKNGVEGVVSTATOSFLATCNGVCMWTFHAGAGSKTLAPGR 1094
 Db 181 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLYLVTRADVIFVRRRGDSRGLLS 240
 1095 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLYLVTRADVIFVRRRGDSRGLLS 1154
 Db 241 PRPVSYLKSSGSGPLCPSGHAGVIFRAAVCTRGVAAYVPIFESMETMR 292
 1155 PRPVSYLKSSGSGPLCPSGHAGVIFRAAVCTRGVAAYVPIFESMETMR 1206

RESULT 13
 ID 080793 PRELIMINARY; PRT: 3010 AA.
 AC 080793
 DT 01-JUN-2003 (TREMBLrel. 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MILE;
 RX MEDLINE=22047193; PubMed=12051758;
 RA Kishine H., Sugiyama K., Hijioka M., Kato N., Takahashi H., Noshi T.,
 RA Nio Y., Hosaka M., Miyazaki Y., Shimotohno K.;
 RT "Subgenomic replicon derived from a cell line infected with the
 RT Hepatitis C virus."
 RL Biochem. Biophys. Res. Commun. 293:993-999(2002).
 DR EMBL: AB040299; BAC54896.1; -;
 DR GO: GO:0019028; C: viral capsid; IEA.
 DR GO: GO:0019031; C: viral envelope; IEA.
 DR GO: GO:0005524; F: ATP binding; IEA.
 DR GO: GO:0008026; F: ATP dependent helicase activity; IEA.
 DR GO: GO:0005489; F: electron transporter activity; IEA.
 DR GO: GO:0003723; F: RNA binding; IEA.
 DR GO: GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0008286; F: serine-type peptidase activity; IEA.
 DR GO: GO:0005198; F: structural molecule activity; IEA.
 DR GO: GO:0016740; F: transferase activity; IEA.
 DR GO: GO:0006118; P: proteolysis and peptidolysis; IEA.
 DR GO: GO:0006350; P: transcription; IEA.
 DR GO: GO:0019079; P: viral genome replication; IEA.
 DR GO: GO:0019087; P: viral transformation; IEA.
 DR InterPro: IPR003003; Cys_Ser_lysine.
 DR InterPro: IPR003451; CysC_heme_BS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV capsid.
 DR InterPro: IPR002521; HCV core.
 DR InterPro: IPR002519; HCV env.
 DR InterPro: IPR002531; HCV NS1.
 DR InterPro: IPR002518; HCV NS2.
 DR InterPro: IPR000745; HCV NS4a.
 DR InterPro: IPR001490; HCV NS4b.
 DR InterPro: IPR002166; HCV NS5a.
 DR InterPro: IPR001650; HCV RdRp.
 DR InterPro: IPR004109; peptidase C29.
 DR InterPro: IPR007095; RNA_pol_DS_P8.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV capsid; 1.
 DR Pfam: PF01542; HCV core; 1.
 DR Pfam: PF01539; HCV env; 1.
 DR Pfam: PF01560; HCV NS1; 1.
 DR Pfam: PF01538; HCV NS2; 1.
 DR Pfam: PF02907; HCV NS3; 1.
 DR Pfam: PF01006; HCV NS4a; 1.
 DR Pfam: PF01001; HCV NS4b; 1.
 DR Pfam: PF01506; HCV NS5a; 1.
 DR Pfam: PF00271; helicase C; 1.
 DR Pfam: PF00998; viral RdRp; 1.
 DR Pfam: PF01662; HCV NS1; 1.
 DR Pfam: PF01662; HCV NS1; 1.
 DR SMART; SMO0487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW Polyprotein.
 SQ SEQUENCE 3010 AA; 327097 MW; EB6418C7A723B686 CRC64;

Query Match 96.9%; Score 1485; DB 12; Length 3010;
 Best Local Similarity 96.2%; Pred. No. 2.3e-123;

Matches 281; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLRACMLVYKRAAGHYVQMAFKALALGTGYVYDHLPLQDMAHAGRLDAVAVEPV 60

DB 915 AAGLRACMLVYKRAAGHYVQMAFKALALGTGYVYDHLPLQDMAHAGRLDAVAVEPV 974

QY 61 IFSDMEVKIITWGADTAAAGDIISGLPVSARGREILLGPADNFEQGMRLAPITAYSQ 120

DB 975 VFSDEMTKIIITWGADTAAAGDIISGLPVSARGREILLGPADNFEQGMRLAPITAYSQ 1034

QY 121 QTRRLGCIITSLTGRDKNQVGEVQVSTATQSFLLATCNGVCMWTFHGAGSKTLAGPK 180

DB 1035 QTRRLGCIITSLTGRDKNQVGEVQVSTATQSFLLATCNGVCMWTFHGAGSKTLAGPK 1094

QY 181 GPITQMTYNVDQDLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPRRRGDSRGLS 240

DB 1095 GPITQMTYNVDQDLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPRRRGDSRGLS 1154

QY 241 PRPVSYLKGSSGGLPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTM 292

DB 1155 PRPVSYLKGSSGGLPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTM 1206

RESULT 14

Q9J311 PRELIMINARY; PRT; 3010 AA.

ID Q9J311

AC Q9J311

DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MD11.

RA Nagaya K., Kurosaki M., Enomoto N., Miyasaka Y., Maruno F., Sato C.;

RT "Characteristics of hepatitis C viral genome associated with disease progression."

RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.

-1- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND RNA (BY SIMILARITY).

DR EMBL; AF207752; AA65942.1; -.

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; PS0329; PS0329.

DR HSRP; P26663; LUXP.

DR GO; GO:0016021; C: integral to membrane; IEA.

DR GO; GO:0019028; C: viral capsid; IEA.

DR GO; GO:0019031; C: viral envelope; IEA.

DR GO; GO:0005524; F: ATP binding; IEA.

DR GO; GO:0008026; F: ATP dependent helicase activity; IEA.

DR GO; GO:0005489; F: electron transporter activity; IEA.

DR GO; GO:0003723; F: RNA binding; IEA.

DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.

DR GO; GO:0008286; F: serine-type peptidase activity; IEA.

DR GO; GO:0005198; F: structural molecule activity; IEA.

DR GO; GO:0016740; F: transferase activity; IEA.

DR GO; GO:0006118; P: electron transport; IEA.

DR GO; GO:0006508; P: proteolysis and peptidolysis; IEA.

DR GO; GO:0006350; P: translation; IEA.

DR GO; GO:0019079; P: viral genome replication; IEA.

DR GO; GO:0019087; P: viral transformation; IEA.

DR InterPro; IPR009003; Cys_ser_lypsin.

DR InterPro; IPR000345; CytC_heme_BS.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV_core.

DR InterPro; IPR002519; HCV_env.

DR InterPro; IPR002531; HCV NS1.

DR InterPro; IPR002518; HCV NS2.

DR InterPro; IPR000745; HCV NS4a.

DR InterPro; IPR001490; HCV NS4b.

DR InterPro; IPR002868; HCV NS5a.

DR InterPro; IPR002166; HCV NS5a.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR004109; Helicase_C29.

DR InterPro; IPR007095; RNA_pol_DS_PS.

DR InterPro; IPR007094; RNA_pol_PSVir.

DR Pfam; PF01543; HCV_capsid; 1.

DR Pfam; PF01542; HCV_core; 1.

DR Pfam; PF01539; HCV_env; 1.

DR Pfam; PF01560; HCV NS1; 1.

DR Pfam; PF01538; HCV NS2; 1.

DR Pfam; PF02907; HCV NS3; 1.

DR Pfam; PF01006; HCV NS4a; 1.

DR Pfam; PF01001; HCV NS4b; 1.

DR Pfam; PF01506; HCV NS5a; 1.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00998; Viral_RdRp; 1.

DR ProDom; PD186062; HCV NS1; 1.

DR SMART; SM00487; DEXDC; 1.

DR PROSITE; PS00190; CYTOCHROME C; 1.

KW Coat protein, RNA-directed RNA polymerase, Glycoprotein, Nonstructural protein, Polyprotein, RNA-directed RNA polymerase, Transferase, Transmembrane.

SQ SEQUENCE 3010 AA; 327193 MW; F88BA81174E19444 CRC64;

Query Match 96.7%; Score 1482; DB 12; Length 3010;

Best Local Similarity 95.5%; Pred. No. 4.2e-123;

Matches 279; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLRACMLVYKRAAGHYVQMAFKALALGTGYVYDHLPLQDMAHAGRLDAVAVEPV 60

DB 915 AAGLRACMLVYKRAAGHYVQMAFKALALGTGYVYDHLPLQDMAHAGRLDAVAVEPV 974

QY 61 IFSDMEVKIITWGADTAAAGDIISGLPVSARGREILLGPADNFEQGMRLAPITAYSQ 120

DB 975 VFSDEMTKIIITWGADTAAAGDIISGLPVSARGREILLGPADNFEQGMRLAPITAYSQ 1034

QY 121 QTRRLGCIITSLTGRDKNQVGEVQVSTATQSFLLATCNGVCMWTFHGAGSKTLAGPK 180

DB 1035 QTRRLGCIITSLTGRDKNQVGEVQVSTATQSFLLATCNGVCMWTFHGAGSKTLAGPK 1094

QY 181 GPITQMTYNVDQDLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPRRRGDSRGLS 240

DB 1095 GPITQMTYNVDQDLVGMQAPPGARSMTPTCGSSDLVLTNRADVIPRRRGDSRGLS 1154

QY 241 PRPVSYLKGSSGGLPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTM 292

DB 1155 PRPVSYLKGSSGGLPLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTM 1206

RESULT 15

O70822 PRELIMINARY; PRT; 361 AA.

ID O70822

AC O70822

DT 01-AUG-1998 (TREMblrel. 07, Created)

DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Polyprotein (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI_TaxID=11103;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98321154; PubMed=9656998;

RA Yamada K., Mori A., Seki M., Kimura T., Yuasa S., Matsuura Y.,

RT Miyamura T.;

RL "Critical point mutations for hepatitis C virus NS3 proteinase.";

RV Virology 246:104-112(1998).

RN [2]

RP SEQUENCE FROM N.A.
 RA Mori A., Yamada K., Kimura J., Koide T., Yuasa S., Yamada E.,
 RA Miyamura T.,
 RT "Enzymatic Characterization of purified NS3 serine proteinase of
 RT hepatitis C virus expressed in Escherichia coli.",
 RL FEBS Lett. 378:37-42(1998).
 DR EMBL: AB013627; BAA28505.1; -.
 DR HSSP: P27958; 1A1V.
 DR GO: GO:0008236; F:serine-type peptidase activity; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO: GO:0019087; P:viral transformation; IEA.
 DR InterPro: IPR009003; Cys_ser_trypsin.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; peptidase_C29.
 DR Pfam: PF01518; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 361 AA; 38144 MW; FE848FF746A383FC CRC64;

Query Match 96.7%; Score 1481; DB 12; Length 361;
 Best local similarity 95.2%; Pred. No. 3, 1e-124;
 Matches 278; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 AQLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAGLRDIAVAPEVY 60
 DB 16 AQLIRACMLVRKAAGHYVQMAFMKLAALTGYVYDHLTPLODMAHAGLRDIAVAPEVY 75
 QY 61 IFSMEVKITWGDPTACGDIIISGLPVSARRGRELLGRADNFEQGMRLAPITAYSQ 120
 DB 76 VFSAMETKVTWGDPTACGDIIISGLPVSARRGRELLGRADNFEQGMRLAPITAYSQ 135
 QY 121 QTRLLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCTVFGAGSKTLA GPR 180
 DB 136 QTRLLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCTVFGAGSKTLA GPR 195
 QY 181 GPITQMTTNVDQDVGQAPPGARSMTPTCTGSSDLYLVTRHADVLPVRRGDSRGSLLS 240
 DB 196 GPITQMTTNVDQDVGQAPPGARSMTPTCTGSSDLYLVTRHADVLPVRRGDSRGSLLS 255
 QY 241 PRPVSYLKSGSGGFLCPSGHGVGIFRAAVCTRGVAKAVDFIPVESMETMR 292
 DB 256 PRPVSYLKSGSGGFLCPSGHGVGIFRAAVCTRGVAKAVDFIPVESMETMR 307

Search completed: May 6, 2004, 09:35:50
 Job time : 28.3102 secs

Fri May 7 13:37:06 2004

us-10-650-585-15.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 11.4605 Seconds
(without alignments)
1315.364 Million cell updates/sec

Title: US-10-650-585-15
Perfect score: 1532
Sequence: 1 AGLIRACMLVRKAGGHV.....RGVAKAVDFIPVSMETMR 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCMTUS COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1473	96.1	1692	3	US-09-263-933-4 Sequence 4, Appl
2	1473	96.1	1692	4	US-09-919-901-4 Sequence 4, Appl
3	1473	96.1	2201	4	US-09-539-601-6 Sequence 6, Appl
4	1473	96.1	2201	4	US-09-539-601-15 Sequence 15, Appl
5	1473	96.1	2307	3	US-08-263-933-2 Sequence 2, Appl
6	1473	96.1	2307	4	US-09-919-901-2 Sequence 2, Appl
7	1473	96.1	3010	4	US-09-539-601-3 Sequence 3, Appl
8	1473	96.1	3010	4	US-09-539-601-21 Sequence 21, Appl
9	1473	96.1	3010	4	US-09-539-601-27 Sequence 27, Appl
10	1470	96.0	1692	3	US-09-263-933-11 Sequence 11, Appl
11	1470	96.0	2307	3	US-08-263-933-9 Sequence 9, Appl
12	1470	96.0	2307	4	US-09-919-901-9 Sequence 9, Appl
13	1466	95.7	3010	4	US-09-539-601-33 Sequence 33, Appl
14	1466	95.4	1692	3	US-09-263-933-18 Sequence 18, Appl
15	1461	95.4	1692	3	US-09-919-901-16 Sequence 16, Appl
16	1461	95.4	2307	3	US-09-263-933-16 Sequence 16, Appl
17	1461	95.4	2307	4	US-09-919-901-16 Sequence 16, Appl
18	1461	95.4	2307	4	US-09-919-901-16 Sequence 16, Appl
19	1450	94.6	3010	3	US-09-014-416-3 Sequence 3, Appl
20	1424	93.0	2013	1	US-08-324-977-12 Sequence 12, Appl
21	1424	93.0	2013	2	US-08-384-616-12 Sequence 12, Appl
22	1424	93.0	2013	2	US-08-904-686A-12 Sequence 12, Appl
23	1424	93.0	2013	3	US-09-315-850-12 Sequence 12, Appl
24	1424	93.0	2201	4	US-08-952-981A-2 Sequence 2, Appl
25	1424	93.0	2620	1	US-08-324-977-32 Sequence 32, Appl
26	1424	93.0	2620	2	US-08-384-616-32 Sequence 32, Appl
27	1424	93.0	2620	2	US-08-904-686A-32 Sequence 32, Appl

28	1424	93.0	2620	3	US-09-315-850-32 Sequence 32, Appl
29	1424	93.0	2621	1	US-08-324-977-36 Sequence 36, Appl
30	1424	93.0	2621	2	US-08-384-616-36 Sequence 36, Appl
31	1424	93.0	2621	2	US-08-904-686A-36 Sequence 36, Appl
32	1424	93.0	2621	3	US-09-315-850-36 Sequence 36, Appl
33	1424	93.0	3010	1	US-08-324-977-12 Sequence 12, Appl
34	1424	93.0	3010	2	US-08-384-616-12 Sequence 12, Appl
35	1424	93.0	3010	2	US-08-384-616-14 Sequence 14, Appl
36	1424	93.0	3010	2	US-08-904-686A-2 Sequence 2, Appl
37	1424	93.0	3010	2	US-08-904-686A-14 Sequence 14, Appl
38	1424	93.0	3010	3	US-09-315-850-2 Sequence 2, Appl
39	1424	93.0	3010	3	US-09-315-850-14 Sequence 14, Appl
40	1362	88.9	3012	3	US-08-811-566-2 Sequence 2, Appl
41	1362	88.9	3012	4	US-09-034-756-2 Sequence 2, Appl
42	1362	88.7	1648	1	US-08-188-281B-12 Sequence 12, Appl
43	1359	88.7	1648	5	PCT-US94-07280-12 Sequence 12, Appl
44	1359	88.7	1648	5	PCT-US95-01067-12 Sequence 12, Appl
45	1359	88.7	1648	5	PCT-US95-01067-12 Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-263-933-4
Sequence 4, Application US/09263933
Patent No. 6280940
GENERAL INFORMATION:
APPLICANT: Potts, Karen B.
APPLICANT: Jackson, Roberta L.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/263,933
CURRENT FILING DATE: 1999-03-08
EARLIER APPLICATION NUMBER: 09/129,611
EARLIER FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
US-09-263-933-4

Query Match 96.1%; Score 1473; DB 3; Length 1692;
Best Local Similarity 94.5%; Pred. No. 8.4e-139;
Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY	1	ACGLIRACMLVRKAGGHVQMAFMKLAALGTGVYDHLTPQDMAHGLRLAVAVEPV	60
DB	194	AGGLIRACMLVRKAGGHVQMAFMKLAALGTGVYDHLTPQDMAHGLRLAVAVEPV	253
QY	61	IFSDMEVKIITGADTAAAGDIIISGLPVARSRGREILIGPADNFEQGRLLAPITAYSQ	120
DB	254	VSDMETKIIITGADTAAAGDIIISGLPVARSRGREILIGPADNFEQGRLLAPITAYSQ	313
QY	121	QTRGLIGCIITSLTGRDKNQVGEVQVSTAQSFATVNVGCVTFVFGASCKTLAGRK	160
DB	314	QTRGLIGCIITSLTGRDKNQVGEVQVSTAQSFATVNVGCVTFVFGASCKTLAGRK	373
QY	181	GPITQWYTWDDLVGMQAPPGARSMTPTCCSSDLVLTFRADVIPIVARRDSRGLSL	240
DB	374	GPITQWYTWDDLVGMQAPPGARSMTPTCCSSDLVLTFRADVIPIVARRDSRGLSL	433
QY	241	PRVSYLKSSGGPILCPGSHAIVGIFRAAVCTRGVAKAVDFIPVSMETMR	292
DB	434	PRVSYLKSSGGPILCPGSHAIVGIFRAAVCTRGVAKAVDFIPVSMETMR	485

RESULT 2
US-09-919-901-4

Sequence 4, Application US/09919901
Patent No. 6599738
GENERAL INFORMATION:
APPLICANT: Potes, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/919,901
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-02-08
PRIOR APPLICATION NUMBER: 09/129,611
PRIOR FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: :
US-09-919-901-4

Query Match 96.1%; Score 1473; DB 4; Length 1692;
Best Local Similarity 94.5%; Pred. No. 8,4e-139;
Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVDHLPLDPMHAGLRDLAVAEV 60
DB 194 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVDHLPLDPMHAGLRDLAVAEV 253
QY 61 IFSDMEVKIITWGDPTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 120
DB 254 VFSDMETKVIITWGDPTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 313
QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPK 180
DB 314 QTRGLGCIITSLTGRDNQVEGEVYVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPK 373
QY 181 GPITQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIVRRRGDSRGSLLS 240
DB 374 GPITQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIVRRRGDSRGSLLS 433
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 292
DB 434 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 485

RESULT 3
US-09-539-601-6
Sequence 6, Application US/09539601C
Patent No. 6630343
GENERAL INFORMATION:
APPLICANT: Bartenschlager, Ralf FW
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
FILE REFERENCE: all sequences
CURRENT APPLICATION NUMBER: US/09/539,601C
CURRENT FILING DATE: 2001-08-30
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
EARLIER FILING DATE: 1999-04-03
NUMBER OF SEQ ID NOS: 51
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 2201
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-539-601-6

Query Match 96.1%; Score 1473; DB 4; Length 2201;
Best Local Similarity 94.9%; Pred. No. 1.2e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVDHLPLDPMHAGLRDLAVAEV 60
DB 106 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVDHLPLDPMHAGLRDLAVAEV 165
QY 61 IFSDMEVKIITWGDPTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 120
DB 166 VFSDMETKVIITWGDPTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 225
QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPK 180
DB 226 QTRGLGCIITSLTGRDNQVEGEVYVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPK 285
QY 181 GPITQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIVRRRGDSRGSLLS 240
DB 286 GPITQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIVRRRGDSRGSLLS 345
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 292
DB 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 397

RESULT 4
US-09-539-601-15
Sequence 15, Application US/09539601C
Patent No. 6630343
GENERAL INFORMATION:
APPLICANT: Bartenschlager, Ralf FW
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
FILE REFERENCE: all sequences
CURRENT APPLICATION NUMBER: US/09/539,601C
CURRENT FILING DATE: 2001-08-30
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
EARLIER FILING DATE: 1999-04-03
NUMBER OF SEQ ID NOS: 51
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 15
LENGTH: 2201
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-539-601-15

Query Match 96.1%; Score 1473; DB 4; Length 2201;
Best Local Similarity 94.9%; Pred. No. 1.2e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVDHLPLDPMHAGLRDLAVAEV 60
DB 106 AAGLIRACMLVRKXAGHYVOMAFMKLAALTGTYYVDHLPLDPMHAGLRDLAVAEV 165
QY 61 IFSDMEVKIITWGDPTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 120
DB 166 VFSDMETKVIITWGDPTAACGDIISGLPVSARGRREILGPADNFEQGMRLAPITAYSQ 225
QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPK 180
DB 226 QTRGLGCIITSLTGRDNQVEGEVYVSTATQSFPLATCVNGVCMVTFHAGSKTLAGPK 285
QY 181 GPITQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIVRRRGDSRGSLLS 240
DB 286 GPITQMTNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIVRRRGDSRGSLLS 345
QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 292
DB 346 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFVPSMETTMR 397

RESULT 5
US-09-263-933-2
Sequence 2, Application US/09263933
Patent No. 6280940
GENERAL INFORMATION:
APPLICANT: Potes, Karen E.

APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/263,933
 CURRENT FILING DATE: 1999-03-08
 EARLIER APPLICATION NUMBER: 09/129,611
 EARLIER FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 2
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 US-09-263-933-2

Query Match 96.1%; Score 1473; DB 3; Length 2307;
 Best Local Similarity 94.5%; Pred. No. 1.3e-138;
 Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLRACMLVRKAAGHYVQAFMKLAALGTYYVDHLPLODMAHAGLDIAVAPEV 60
 DB 286 AAGLRACMLVRKAAGHYVQAFMKLAALGTYYVDHLPLODMAHAGLDIAVAPEV 345
 QY 61 IFSDMEVKITMGADTAACGDIISGLPVASARRGEIILGPADNPEQGMRLAPITAYSQ 120
 DB 346 VPSDMETKITMGADTAACGDIISGLPVASARRGEIILGPADNPEQGMRLAPITAYSQ 405
 QY 121 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCNGVCMTVPHGASKTLAGPK 180
 DB 406 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCNGVCMTVPHGASKTLAGPK 465
 QY 181 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRGDSRGLLS 240
 DB 466 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRGDSRGLLS 525
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292
 DB 526 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 577

RESULT 6
 US-09-919-901-2
 Sequence 2, Application US/09919901
 Patent No. 6599738
 GENERAL INFORMATION:
 APPLICANT: Potts, Karen E.
 APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/919,901
 CURRENT FILING DATE: 2001-08-02
 PRIOR APPLICATION NUMBER: 09/263,933
 PRIOR FILING DATE: 1999-02-08
 PRIOR APPLICATION NUMBER: 09/129,611
 PRIOR FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 2
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: :
 US-09-919-901-2

Query Match 96.1%; Score 1473; DB 4; Length 2307;
 Best Local Similarity 94.5%; Pred. No. 1.3e-138;
 Matches 276; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLRACMLVRKAAGHYVQAFMKLAALGTYYVDHLPLODMAHAGLDIAVAPEV 60
 DB 286 AAGLRACMLVRKAAGHYVQAFMKLAALGTYYVDHLPLODMAHAGLDIAVAPEV 345
 QY 61 IFSDMEVKITMGADTAACGDIISGLPVASARRGEIILGPADNPEQGMRLAPITAYSQ 120
 DB 346 VPSDMETKITMGADTAACGDIISGLPVASARRGEIILGPADNPEQGMRLAPITAYSQ 405
 QY 121 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCNGVCMTVPHGASKTLAGPK 180
 DB 406 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCNGVCMTVPHGASKTLAGPK 465
 QY 181 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRGDSRGLLS 240
 DB 466 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRGDSRGLLS 525
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292
 DB 526 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 577

RESULT 7
 US-09-539-601-3
 Sequence 3, Application US/09539601C
 Patent No. 6630343
 GENERAL INFORMATION:
 APPLICANT: Bartschlagel, Ralf FW
 TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 FILE REFERENCE: all sequences
 CURRENT APPLICATION NUMBER: US/09/539,601C
 CURRENT FILING DATE: 2001-08-30
 EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
 EARLIER FILING DATE: 1999-04-03
 NUMBER OF SEQ ID NOS: 51
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO 3
 LENGTH: 3010
 TYPE: PRT
 ORGANISM: Hepatitis C virus
 US-09-539-601-3

Query Match 96.1%; Score 1473; DB 4; Length 3010;
 Best Local Similarity 94.9%; Pred. No. 1.9e-138;
 Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLRACMLVRKAAGHYVQAFMKLAALGTYYVDHLPLODMAHAGLDIAVAPEV 60
 DB 915 AAGLRACMLVRKAAGHYVQAFMKLAALGTYYVDHLPLODMAHAGLDIAVAPEV 974
 QY 61 IFSDMEVKITMGADTAACGDIISGLPVASARRGEIILGPADNPEQGMRLAPITAYSQ 120
 DB 975 VPSDMETKITMGADTAACGDIISGLPVASARRGEIILGPADNPEQGMRLAPITAYSQ 1034
 QY 121 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCNGVCMTVPHGASKTLAGPK 180
 DB 1035 QTRGLGCIITSLTRDKNOVEGEVQVSTATQSFATCNGVCMTVPHGASKTLAGPK 1094
 QY 181 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRGDSRGLLS 240
 DB 1095 GPITQMTNVDDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRGDSRGLLS 1154
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 292
 DB 1155 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIVESMETTMR 1206

RESULT 8
 US-09-539-601-21
 Sequence 21, Application US/09539601C
 Patent No. 6630343
 GENERAL INFORMATION:
 APPLICANT: Bartschlagel, Ralf FW
 TITLE OF INVENTION: Hepatitis C Virus Cell Culture System

FILE REFERENCE: all sequences
CURRENT APPLICATION NUMBER: US/09/539,601C
CURRENT FILING DATE: 2001-08-30
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
EARLIER FILING DATE: 1999-04-03
NUMBER OF SEQ ID NOS: 51
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 21
LENGTH: 3010
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-539-601-21

Query Match 96.1%; Score 1473; DB 4; Length 3010;
Best Local Similarity 94.9%; Pred. No. 1.9e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 60
DB 915 AHGILIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 974
QY 61 IFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 120
DB 975 VFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 1094
QY 181 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVIPIVRRGDSRGLLS 240
DB 1095 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVIPIVRRGDSRGLLS 1154
QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 292
DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 1206

RESULT 9
US-09-539-601-27
Sequence 27, Application US/09539601C
Patent No. 6630343
GENERAL INFORMATION:
APPLICANT: Barteneschlager, Ralf FW
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
FILE REFERENCE: all sequences
CURRENT APPLICATION NUMBER: US/09/539,601C
CURRENT FILING DATE: 2001-08-30
EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
EARLIER FILING DATE: 1999-04-03
NUMBER OF SEQ ID NOS: 51
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 27
LENGTH: 3010
TYPE: PRT
ORGANISM: Hepatitis C virus
US-09-539-601-27

Query Match 96.1%; Score 1473; DB 4; Length 3010;
Best Local Similarity 94.9%; Pred. No. 1.9e-138;
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGLIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 60
DB 915 AHGILIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 974
QY 61 IFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 120
DB 975 VFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 1094

QY 181 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVIPIVRRGDSRGLLS 240
DB 1095 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVIPIVRRGDSRGLLS 1154
QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 292
DB 1155 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 1206

RESULT 10
US-09-263-933-11
Sequence 11, Application US/09263933
Patent No. 6280940
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/263,933
CURRENT FILING DATE: 1999-03-08
EARLIER APPLICATION NUMBER: 09/129,611
EARLIER FILING DATE: 1998-08-05
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 11
LENGTH: 1692
TYPE: PRT
ORGANISM: Artificial Sequence
US-09-263-933-11

Query Match 96.0%; Score 1470; DB 3; Length 1692;
Best Local Similarity 94.2%; Pred. No. 1.7e-118;
Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 60
DB 194 AAGLIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 253
QY 61 IFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 120
DB 254 VFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 313
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 314 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 373
QY 181 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVIPIVRRGDSRGLLS 240
DB 374 GPITQMTNNVDODLVGQAPPGARSMTPTCGSSDLYLTRHADVIPIVRRGDSRGLLS 433
QY 241 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 292
DB 434 PRPVSYLKGSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTR 485

RESULT 11
US-09-919-901-11
Sequence 11, Application US/0919901
Patent No. 6599738
GENERAL INFORMATION:
APPLICANT: Potts, Karen E.
APPLICANT: Jackson, Roberta L.
APPLICANT: Patrick, Amy K.
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
FILE REFERENCE: 0125-0005A
CURRENT APPLICATION NUMBER: US/09/919,901
CURRENT FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 09/263,933
PRIOR FILING DATE: 1999-02-08

QY 1 AAGLIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 60
DB 915 AHGILIRACMIVKRAAGHYVQMAFMKLAALTGYVVDHLTPLODMAHAGRLDAVAPEV 974
QY 61 IFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 120
DB 975 VFSDEMEKIIITWGDPTAACGDIISGLPVSARRGREILGPDNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCMVTFHGAGSKTLAGPK 1094

PRIOR APPLICATION NUMBER: 09/129,611
 PRIOR FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 11
 LENGTH: 1692
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: :
 US-09-919-901-11

Query Match 96.0%; Score 1470; DB 4; Length 1692;
 Best Local Similarity 94.2%; Pred. No. 1,7e-138;
 Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAPEV 60
 DB 194 AAGLIHACMLVRKAGHYVQMAFMKLGALTGTYYNHLPLRDMAHAGRLDAVAPEV 253
 QY 61 IFSDMEVKIIITWGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 120
 DB 254 VFSMETERKIITWGADTAACGDIILGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 313
 QY 121 QTRGLGCIITSLTRGRDKNOVEGVVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180
 DB 314 QTRGLGCIITSLTRGRDKNOVEGVVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 373
 QY 181 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGSLLS 240
 DB 374 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGSLLS 433
 QY 241 PRPVSYLKSGSGGPIILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETMR 292
 DB 434 PRPVSYLKSGSGGPIILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETMR 485

RESULT 12

US-09-263-933-9
 Sequence 9, Application US/09263933
 Patent No. 6280940
 GENERAL INFORMATION:
 APPLICANT: Potts, Karen E.
 APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/263,933
 CURRENT FILING DATE: 1999-03-08
 EARLIER APPLICATION NUMBER: 09/129,611
 EARLIER FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 9
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 US-09-263-933-9

Query Match 96.0%; Score 1470; DB 3; Length 2307;
 Best Local Similarity 94.2%; Pred. No. 2.6e-138;
 Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAPEV 60
 DB 286 AAGLIHACMLVRKAGHYVQMAFMKLGALTGTYYNHLPLRDMAHAGRLDAVAPEV 345
 QY 61 IFSDMEVKIIITWGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 120
 DB 346 VFSMETERKIITWGADTAACGDIILGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 405
 QY 121 QTRGLGCIITSLTRGRDKNOVEGVVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180

DB 406 QTRGLGCIITSLTRGRDKNOVEGVVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 465
 QY 181 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGSLLS 240
 DB 466 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGSLLS 525
 QY 241 PRPVSYLKSGSGGPIILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETMR 292
 DB 526 PRPVSYLKSGSGGPIILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETMR 577

RESULT 13

US-09-919-901-9
 Sequence 9, Application US/09919901
 Patent No. 6539738
 GENERAL INFORMATION:
 APPLICANT: Potts, Karen E.
 APPLICANT: Jackson, Roberta L.
 APPLICANT: Patrick, Amy K.
 TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
 TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
 FILE REFERENCE: 0125-0005A
 CURRENT APPLICATION NUMBER: US/09/919,901
 CURRENT FILING DATE: 2001-08-02
 PRIOR APPLICATION NUMBER: 09/263,933
 PRIOR FILING DATE: 1999-02-08
 PRIOR FILING DATE: 1998-08-05
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 9
 LENGTH: 2307
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: :
 US-09-919-901-9

Query Match 96.0%; Score 1470; DB 4; Length 2307;
 Best Local Similarity 94.2%; Pred. No. 2.6e-138;
 Matches 275; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGHYVQMAFMKLAALTGTYYDHLTPLODMAHAGRLDAVAPEV 60
 DB 286 AAGLIHACMLVRKAGHYVQMAFMKLGALTGTYYNHLPLRDMAHAGRLDAVAPEV 345
 QY 61 IFSDMEVKIIITWGADTAACGDIISGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 120
 DB 346 VFSMETERKIITWGADTAACGDIILGLPVSARGREIILGPADNFEQGMRLAPITAYSQ 405
 QY 121 QTRGLGCIITSLTRGRDKNOVEGVVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 180
 DB 406 QTRGLGCIITSLTRGRDKNOVEGVVSTATOSFLATCVNGVCMVTFHAGSKTLAEPK 465
 QY 181 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGSLLS 240
 DB 466 GPITQWYTNVDQDLVGMQAPPGARSMTPTCGSSDLVYTRHADVIPIVRRRDSRGSLLS 525
 QY 241 PRPVSYLKSGSGGPIILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETMR 292
 DB 526 PRPVSYLKSGSGGPIILCPSGHAGVIFRAAVCTRGVAKAVDFIVPESMETMR 577

RESULT 14

US-09-539-601-33
 Sequence 33, Application US/09539601C
 Patent No. 6630343
 GENERAL INFORMATION:
 APPLICANT: Bartenschlager, Ralf FW
 TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
 FILE REFERENCE: all sequences
 CURRENT APPLICATION NUMBER: US/09/539,601C.

```

; CURRENT FILING DATE: 2001-08-30
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; EARLIER FILING DATE: 1999-04-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 3010
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-09-539-601-33

```

```

Query Match          95.7%; Score 1466; DB 4; Length 3010;
Best Local Similarity 94.5%; Pred. No. 9,7e-138;
Matches 276; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 1 AAGLIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAAVAVPEV 60
DB 915 AHGILIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAAVAVPEV 974
QY 61 IFSDMETKVTITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 120
DB 975 VFSDMETKVTITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 1034
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMTVPHGAGSKTLAGPK 180
DB 1035 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMTVPHGAGSKTLAGPK 1094
QY 181 GPITOMTYNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 240
DB 1095 GPITOMTYNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 1154
QY 241 PRPVSYLKSGSGGFLCPSGHAAGIFPRAAVCTRGVAKAVDFPVESMETTMR 292
DB 1155 PRPVSYLKSGSGGFLCPSGHAAGIFPRAAVCTRGVAKAVDFPVESMETTMR 1206

```

```

RESULT 15
US-09-263-933-18
; Sequence 18, Application US/09263933
; Patent No. 6280940
; GENERAL INFORMATION:
; APPLICANT: Potes, Karen E.
; APPLICANT: Jackson, Roberta L.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/09/263,933
; EARLIER FILING DATE: 1999-03-08
; EARLIER APPLICATION NUMBER: 09/129,611
; EARLIER FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-263-933-18

```

```

Query Match          95.4%; Score 1461; DB 3; Length 1692;
Best Local Similarity 93.8%; Pred. No. 1.4e-137;
Matches 274; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 1 AAGLIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAAVAVPEV 60
DB 194 AAGLIRACMLVRKAGGHYQMAFMKLAALTGYVDHLTPLODMAHAGLRDLAAVAVPEV 253
QY 61 IFSDMETKVTITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 120
DB 254 VFSDMETKVTITWGADTAACGDIISGLPVSARRGREIILGPADNFEQGMRLAPITAYSQ 313
QY 121 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMTVPHGAGSKTLAGPK 180

```

```

DB 314 QTRGLGCIITSLTGRDNQVEGEVQVSTATQSFATCVNGVCMTVPHGAGSKTLAGPK 373
QY 181 GPITOMTYNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 240
DB 374 GPITOMTYNNVDQDLVGMQAPPGARSMTPTCTGSSDLYLVTRHADVIIVRRRGDSRGSLLS 433
QY 241 PRPVSYLKSGSGGFLCPSGHAAGIFPRAAVCTRGVAKAVDFPVESMETTMR 292
DB 434 PRPVSYLKSGSGGFLCPSGHAAGIFPRAAVCTRGVAKAVDFPVESMETTMR 485

```

```

Search completed: May 6, 2004, 09:39:06
Job time : 12.4605 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:30:56 ; Search time 29.8706 Seconds

(without alignments)
2713.357 Million cell updates/sec

Title: US-10-650-585-15

Perfect score: 1532

Sequence: 1 AAGLIRACMLVRRKAGSHV.....RGVAKAVDPFVNSMETTWK 292

Scoring table: BLOSUM62

Gapcp 10.0 , Gapext 0.5

Searched: 1140673 seqs, 27366755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/prodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubppaa/PCR_NEW_PUB.pep:*
3: /cgn2_6/prodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/1/pubppaa/PCRUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/prodata/1/pubppaa/US09_NEW_PUB.pep:*
12: /cgn2_6/prodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/prodata/1/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/1/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/prodata/1/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/prodata/1/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/prodata/1/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1532	100.0	292	13	US-10-017-736-15
2	1532	100.0	292	16	US-10-650-585-15
3	1532	100.0	303	13	US-10-017-736-10
4	1532	100.0	303	16	US-10-650-585-10
5	1532	100.0	334	13	US-10-017-736-4
6	1532	100.0	334	16	US-10-650-585-4
7	1532	100.0	341	16	US-10-017-736-14
8	1532	100.0	341	13	US-10-650-585-14
9	1532	100.0	352	16	US-10-017-736-13
10	1532	100.0	352	13	US-10-650-585-13
11	1532	100.0	380	13	US-10-017-736-12
12	1532	100.0	380	16	US-10-650-585-12
13	1532	100.0	393	13	US-10-017-736-11
14	1532	100.0	393	16	US-10-650-585-11
15	1532	100.0	409	13	US-10-017-736-2

16	1532	100.0	409	16	US-10-650-585-2	Sequence 2, Appli
17	1532	99.4	303	13	US-10-017-736-18	Sequence 18, Appli
18	1532	99.4	303	16	US-10-650-585-18	Sequence 18, Appli
19	1532	99.3	303	13	US-10-017-736-16	Sequence 16, Appli
20	1532	99.3	303	16	US-10-650-585-16	Sequence 16, Appli
21	1532	98.8	301	13	US-10-017-736-17	Sequence 17, Appli
22	1532	98.8	301	16	US-10-650-585-17	Sequence 17, Appli
23	1473	96.1	1992	10	US-09-919-901-4	Sequence 4, Appli
24	1473	96.1	1992	14	US-10-191-966-4	Sequence 4, Appli
25	1473	96.1	2201	13	US-10-029-907-3	Sequence 3, Appli
26	1473	96.1	2201	16	US-10-309-561-3	Sequence 3, Appli
27	1473	96.1	2307	10	US-09-919-901-2	Sequence 2, Appli
28	1473	96.1	2307	14	US-10-191-966-2	Sequence 2, Appli
29	1473	96.1	3010	12	US-10-467-000-1	Sequence 1, Appli
30	1470	96.0	1692	10	US-09-919-901-11	Sequence 11, Appli
31	1470	96.0	1692	14	US-10-191-966-11	Sequence 11, Appli
32	1470	96.0	2307	10	US-09-919-901-9	Sequence 9, Appli
33	1470	96.0	2307	14	US-10-191-966-9	Sequence 9, Appli
34	1461	95.4	1692	10	US-09-919-901-18	Sequence 18, Appli
35	1461	95.4	1692	14	US-10-191-966-18	Sequence 18, Appli
36	1461	95.4	2307	10	US-09-919-901-16	Sequence 16, Appli
37	1461	95.4	2307	14	US-10-191-966-16	Sequence 16, Appli
38	1424	93.0	2201	13	US-10-085-476-2	Sequence 2, Appli
39	1362	88.9	3011	9	US-09-742-659-4	Sequence 4, Appli
40	1362	88.9	3011	10	US-09-891-894-3	Sequence 3, Appli
41	1362	88.9	3011	14	US-10-184-150-3	Sequence 3, Appli
42	1362	88.9	3011	15	US-10-328-997-3	Sequence 3, Appli
43	1362	88.9	3012	9	US-09-238-076-2	Sequence 2, Appli
44	1362	88.9	3012	10	US-09-995-937-2	Sequence 2, Appli
45	1362	88.9	3012	10	US-09-917-563-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-10-017-736-15
; Sequence 15, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 292
; TYPE: PRT
; ORGANISM: HCV
; US-10-017-736-15

Query Match	100.0%	Score 1532	DB 13	Length 292
Best Local Similarity	100.0%	Pred. No. 1e-148		
Matches 292	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	AAGLIRACMLVRRKAGSHVYMAFKKALATGTYVDHLPQDMNAHAGLRDLAAVEEV	60	
DB	1	AAGLIRACMLVRRKAGSHVYMAFKKALATGTYVDHLPQDMNAHAGLRDLAAVEEV	60	
QY	61	ITSDMEVKIITWGADTAAAGDITISGLPVASARRGREITLLGPADNFFGQGRLLIAPITVYSQ	120	
DB	61	ITSDMEVKIITWGADTAAAGDITISGLPVASARRGREITLLGPADNFFGQGRLLIAPITVYSQ	120	
QY	121	QTRGLGCIITLTGSDKNQVGEVYVSTAFATVNVGCMVTFHGAASKTLAPGX	180	
DB	121	QTRGLGCIITLTGSDKNQVGEVYVSTAFATVNVGCMVTFHGAASKTLAPGX	180	
QY	181	GPITQWTVNDVDDLVGWAAPGARSMTPCTCGSSDLVYVTRHADYIPVARRDSSGSLIS	240	

Db 181 GPTQMTNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIFVRRRGRSGSLLS 240

Qy 241 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 292

Db 241 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 292

RESULT 2

US-10-650-585-15

Sequence 15, Application US/10650585

Publication No. US2004007066A1

GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.

TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

FILE REFERENCE: 13/082

CURRENT APPLICATION NUMBER: US/10/650,585

CURRENT FILING DATE: 2003-08-28

PRIOR APPLICATION NUMBER: US/10/017,736A

PRIOR FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: 60/256,031

PRIOR FILING DATE: 2000-12-15

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 15

LENGTH: 292

TYPE: PR

ORGANISM: HCV

US-10-650-585-15

Query Match 100.0%; Score 1532; DB 16; Length 292;

Best Local Similarity 100.0%; Pred. No. 1e-148;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGLRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

Db 1 AAGLRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

Qy 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLLAPITAYSQ 120

Db 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLLAPITAYSQ 120

Qy 121 QTRGLGCIITSLTRGRDNQVEGEVYSTATQSFPLATCNVGVCTVYFHGAGSKTLGPK 180

Db 121 QTRGLGCIITSLTRGRDNQVEGEVYSTATQSFPLATCNVGVCTVYFHGAGSKTLGPK 180

Qy 181 GPTQMTNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIFVRRRGRSGSLLS 240

Db 181 GPTQMTNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIFVRRRGRSGSLLS 240

Qy 241 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 292

Db 241 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 292

RESULT 3

US-10-017-736-10

Sequence 10, Application US/10017736

Publication No. US20020192640A1

GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.

TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

FILE REFERENCE: 13/082

CURRENT APPLICATION NUMBER: US/10/017,736

CURRENT FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: 60/256,031

PRIOR FILING DATE: 2000-12-15

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10

LENGTH: 303

TYPE: PR

ORGANISM: HCV

US-10-017-736-10

Query Match 100.0%; Score 1532; DB 13; Length 303;

Best Local Similarity 100.0%; Pred. No. 1.1e-148;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGLRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

Db 1 AAGLRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 71

Qy 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLLAPITAYSQ 120

Db 72 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLLAPITAYSQ 131

Qy 121 QTRGLGCIITSLTRGRDNQVEGEVYSTATQSFPLATCNVGVCTVYFHGAGSKTLGPK 180

Db 132 QTRGLGCIITSLTRGRDNQVEGEVYSTATQSFPLATCNVGVCTVYFHGAGSKTLGPK 191

Qy 181 GPTQMTNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIFVRRRGRSGSLLS 240

Db 192 GPTQMTNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIFVRRRGRSGSLLS 251

Qy 241 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 292

Db 252 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 303

RESULT 4

US-10-650-585-10

Sequence 10, Application US/10650585

Publication No. US2004007066A1

GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.

TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

FILE REFERENCE: 13/082

CURRENT APPLICATION NUMBER: US/10/650,585

CURRENT FILING DATE: 2003-08-28

PRIOR APPLICATION NUMBER: US/10/017,736A

PRIOR FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: 60/256,031

PRIOR FILING DATE: 2000-12-15

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10

LENGTH: 303

TYPE: PR

ORGANISM: HCV

US-10-650-585-10

Query Match 100.0%; Score 1532; DB 16; Length 303;

Best Local Similarity 100.0%; Pred. No. 1.1e-148;

Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGLRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 60

Db 12 AAGLRACMLVRKAAAGHYVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAEPV 71

Qy 61 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLLAPITAYSQ 120

Db 72 IFSMEVKIITWGDPTAACGDIISGLPVARSREIILGPADNPEGQWRLLAPITAYSQ 131

Qy 121 QTRGLGCIITSLTRGRDNQVEGEVYSTATQSFPLATCNVGVCTVYFHGAGSKTLGPK 180

Db 132 QTRGLGCIITSLTRGRDNQVEGEVYSTATQSFPLATCNVGVCTVYFHGAGSKTLGPK 191

Qy 181 GPTQMTNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIFVRRRGRSGSLLS 240

Db 192 GPTQMTNVDQDLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIFVRRRGRSGSLLS 251

Qy 241 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 292

Db 252 PRPVSYLKSSGGGPGILCPGSHAVGIFRAAVCTRGVAKAVDFIVESNETTMR 303

RESULT 5
 US-10-017-736-4
 ; Sequence 4, Application US/10017736
 ; Publication No. US20020192640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/017,736
 ; CURRENT FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 334
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-017-736-4

Query Match 100.0%; Score 1532; DB 13; Length 334;
 Best Local Similarity 100.0%; Pred. No. 1.2e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 60
 DB 27 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 86
 QY 61 IFSDMVEKIIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 120
 DB 87 IFSDMVEKIIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 146
 QY 121 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSLATCVNGVCWTFVHGAGSKTLAEPK 180
 DB 147 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSLATCVNGVCWTFVHGAGSKTLAEPK 206
 QY 181 GPITQMTNVDDQLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 207 GPITQMTNVDDQLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 266
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
 DB 267 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 318

RESULT 6
 US-10-650-585-4
 ; Sequence 4, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585
 ; CURRENT FILING DATE: 2003-08-28
 ; PRIOR APPLICATION NUMBER: US/10/017,736A
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 334
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-650-585-4

Query Match 100.0%; Score 1532; DB 16; Length 334;
 Best Local Similarity 100.0%; Pred. No. 1.2e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 60

DB 27 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 86
 QY 61 IFSDMVEKIIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 120
 DB 87 IFSDMVEKIIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 146
 QY 121 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSLATCVNGVCWTFVHGAGSKTLAEPK 180
 DB 147 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSLATCVNGVCWTFVHGAGSKTLAEPK 206
 QY 181 GPITQMTNVDDQLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 207 GPITQMTNVDDQLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 266
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
 DB 267 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 318

RESULT 7
 US-10-017-736-14
 ; Sequence 14, Application US/10017736
 ; Publication No. US20020192640A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/017,736
 ; CURRENT FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: 60/256,031
 ; PRIOR FILING DATE: 2000-12-15
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 14
 ; LENGTH: 341
 ; TYPE: PRT
 ; ORGANISM: HCV
 US-10-017-736-14

Query Match 100.0%; Score 1532; DB 13; Length 341;
 Best Local Similarity 100.0%; Pred. No. 1.3e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 60
 DB 50 AAGLIRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGRLDAVAPEV 109
 QY 61 IFSDMVEKIIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 120
 DB 110 IFSDMVEKIIITWGADTAACGDIISGLPVARSRGREILLGPADNFEQGWRLLAPITAYSQ 169
 QY 121 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSLATCVNGVCWTFVHGAGSKTLAEPK 180
 DB 170 QTRGLGCIITSLTRDKNQVEGEVQVSTATQSLATCVNGVCWTFVHGAGSKTLAEPK 229
 QY 181 GPITQMTNVDDQLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 230 GPITQMTNVDDQLVGMQAPPGARSMTPCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 289
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
 DB 290 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341

RESULT 8
 US-10-650-585-14
 ; Sequence 14, Application US/10650585
 ; Publication No. US20040077066A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 ; FILE REFERENCE: 13/082
 ; CURRENT APPLICATION NUMBER: US/10/650,585

CURRENT FILING DATE: 2003-08-28
 PRIOR APPLICATION NUMBER: US/10/017,736A
 PRIOR FILING DATE: 2001-12-14
 PRIOR APPLICATION NUMBER: 60/256,031
 PRIOR FILING DATE: 2000-12-15
 NUMBER OF SEQ ID NOS: 21
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 14
 LENGTH: 341
 TYPE: PRT
 ORGANISM: HCV
 US-10-650-585-14

Query Match 100.0%; Score 1532; DB 16; Length 341;
 Best Local Similarity 100.0%; Pred. No. 1.3e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 60
 DB 50 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 109
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 120
 DB 110 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 169
 QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 170 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 229
 QY 181 GPITQMTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 230 GPITQMTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 289
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
 DB 290 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 341

RESULT 9

US-10-017-736-13
 Sequence 13, Application US/10017736
 Publication No. US20020192640A1
 GENERAL INFORMATION:
 APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 FILE REFERENCE: 13/082
 CURRENT APPLICATION NUMBER: US/10/017,736
 CURRENT FILING DATE: 2001-12-14
 PRIOR APPLICATION NUMBER: 60/256,031
 PRIOR FILING DATE: 2000-12-15
 NUMBER OF SEQ ID NOS: 21
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13
 LENGTH: 352
 TYPE: PRT
 ORGANISM: HCV
 US-10-017-736-13

Query Match 100.0%; Score 1532; DB 13; Length 352;
 Best Local Similarity 100.0%; Pred. No. 1.3e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 60
 DB 61 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 120
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 120
 DB 121 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 180
 QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 181 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 240

QY 181 GPITQMTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 241 GPITQMTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
 DB 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 352

RESULT 10

US-10-650-585-13
 Sequence 13, Application US/10650585
 Publication No. US2004007066A1
 GENERAL INFORMATION:
 APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 FILE REFERENCE: 13/082
 CURRENT APPLICATION NUMBER: US/10/650,585
 CURRENT FILING DATE: 2003-08-28
 PRIOR APPLICATION NUMBER: US/10/017,736A
 PRIOR FILING DATE: 2001-12-14
 PRIOR APPLICATION NUMBER: 60/256,031
 PRIOR FILING DATE: 2000-12-15
 NUMBER OF SEQ ID NOS: 21
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13
 LENGTH: 352
 TYPE: PRT
 ORGANISM: HCV
 US-10-650-585-13

Query Match 100.0%; Score 1532; DB 16; Length 352;
 Best Local Similarity 100.0%; Pred. No. 1.3e-148;
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 60
 DB 61 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 120
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 120
 DB 121 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 180
 QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 181 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 240
 QY 181 GPITQMTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
 DB 241 GPITQMTNVDODLVGMQAPPGARSMTCTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 300
 QY 241 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 292
 DB 301 PRPVSYLKSSGGPILCPSGHAGVIFRAAVCTRGVAKAVDFIPVESMETTMR 352

RESULT 11

US-10-017-736-12
 Sequence 12, Application US/10017736
 Publication No. US20020192640A1
 GENERAL INFORMATION:
 APPLICANT: Boehringer Ingelheim (Canada) Ltd.
 TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
 FILE REFERENCE: 13/082
 CURRENT APPLICATION NUMBER: US/10/017,736
 CURRENT FILING DATE: 2001-12-14
 PRIOR APPLICATION NUMBER: 60/256,031
 PRIOR FILING DATE: 2000-12-15
 NUMBER OF SEQ ID NOS: 21
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12
 LENGTH: 380

QY 1 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 60
 DB 61 AAGGIRACMLVRKKAAGHYVQMAFMKLAALTGTYYVDHLPLQDMAHAGLRDLAVAVEPV 120
 QY 61 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 120
 DB 121 IFSMEVKIITWGADTAACGDIISGLPVSARGREIILGPADNPEGQWRLLAPITAYSQ 180
 QY 121 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 180
 DB 181 QTRGLGCIITSLTGRDNQVEGEVYVSTATOSFLATCVNGVCMVTFHAGSKTLAGPK 240

```
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-12

Query Match      100.0%; Score 1532; DB 13; Length 380;
Best Local Similarity 100.0%; Pred. No. 1,5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGLRDLAAVEPV 60
DB 89 AAGIIRACMLVRKAAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGLRDLAAVEPV 148
QY 61 IFSDMVEKIIITWGDITAAAGDIIISGLPVSARREIILGPADNFEQGWRLLAPITAYSQ 120
DB 149 IFSDMVEKIIITWGDITAAAGDIIISGLPVSARREIILGPADNFEQGWRLLAPITAYSQ 208
QY 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 180
DB 209 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 268
QY 181 GPITQMTNVDDIVGMOAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRGRDSRGLLS 240
DB 269 GPITQMTNVDDIVGMOAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRGRDSRGLLS 328
QY 241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTMR 292
DB 329 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTMR 380

RESULT 12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 380
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-12

Query Match      100.0%; Score 1532; DB 16; Length 380;
Best Local Similarity 100.0%; Pred. No. 1,5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGLRDLAAVEPV 60
DB 89 AAGIIRACMLVRKAAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGLRDLAAVEPV 148
QY 61 IFSDMVEKIIITWGDITAAAGDIIISGLPVSARREIILGPADNFEQGWRLLAPITAYSQ 120
DB 149 IFSDMVEKIIITWGDITAAAGDIIISGLPVSARREIILGPADNFEQGWRLLAPITAYSQ 208
QY 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 180
DB 209 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 268
QY 181 GPITQMTNVDDIVGMOAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRGRDSRGLLS 240
DB 269 GPITQMTNVDDIVGMOAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRGRDSRGLLS 328
QY 241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTMR 292
DB 329 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTMR 292
```

```
DB 329 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTMR 380

RESULT 13
US-10-017-736-11
; Sequence 11, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-11

Query Match      100.0%; Score 1532; DB 13; Length 393;
Best Local Similarity 100.0%; Pred. No. 1,5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGIIRACMLVRKAAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGLRDLAAVEPV 60
DB 102 AAGIIRACMLVRKAAAGHYVOMAFMKLAALTGYVYDHLTPLODMAHAGLRDLAAVEPV 161
QY 61 IFSDMVEKIIITWGDITAAAGDIIISGLPVSARREIILGPADNFEQGWRLLAPITAYSQ 120
DB 162 IFSDMVEKIIITWGDITAAAGDIIISGLPVSARREIILGPADNFEQGWRLLAPITAYSQ 221
QY 121 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 180
DB 222 QTRGLGCIITSLTGRDKNQVEGEVQVSTATQSFATCVNGVCWTFVHGAGSKTLAGPK 281
QY 181 GPITQMTNVDDIVGMOAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRGRDSRGLLS 240
DB 282 GPITQMTNVDDIVGMOAPPGARSMTPTCGSSDLYLVTRHADVIIVRRRGRDSRGLLS 341
QY 241 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTMR 292
DB 342 PRPVSYLKSGSGGPLLCPSGHAGVIFRAAVCTRGVAKAVDFIPVSMETTMR 393

RESULT 14
US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/650,585
; PRIOR FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/10/017,736A
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
US-10-650-585-11

Query Match      100.0%; Score 1532; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 1,5e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AOGILRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPV 60
DB      102 AOGILRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPV 161
QY      61 IFSMEVKIITWGADTAACGDIISGLPVSAARGREILGPADNFEQGWRLLAPITAYSQ 120
DB      162 IFSMEVKIITWGADTAACGDIISGLPVSAARGREILGPADNFEQGWRLLAPITAYSQ 221
QY      121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCWTFHGAGSKTLGPK 180
DB      222 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCWTFHGAGSKTLGPK 281
QY      181 GPITOMTYNVQDILVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
DB      282 GPITOMTYNVQDILVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 341
QY      241 PRPVSYLKSSGGGFLCPSGHAAGIFRAAVCTRGVAKAVDFIPIVESMETMR 292
DB      342 PRPVSYLKSSGGGFLCPSGHAAGIFRAAVCTRGVAKAVDFIPIVESMETMR 393
```

RESULT 15

```
US-10-017-736-2
; Sequence 2, Application US/10017736
; Publication No. US20020192640A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: HCV
US-10-017-736-2
```

```
Query Match      100.0%; Score 1532; DB 13; Length 409;
Best Local Similarity 100.0%; Pred. No. 1.6e-148;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AOGILRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPV 60
DB      106 AOGILRACMLVRKAGGHVQMAFMKLAALTGTYYVDHLTPLODMAHAGLRDLAVAVEPV 165
QY      61 IFSMEVKIITWGADTAACGDIISGLPVSAARGREILGPADNFEQGWRLLAPITAYSQ 120
DB      166 IFSMEVKIITWGADTAACGDIISGLPVSAARGREILGPADNFEQGWRLLAPITAYSQ 225
QY      121 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCWTFHGAGSKTLGPK 180
DB      226 QTRGLGCIITSLTGRDNQVEGEVQVSTATOSFLATCVNGVCWTFHGAGSKTLGPK 285
QY      181 GPITOMTYNVQDILVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 240
DB      286 GPITOMTYNVQDILVGMQAPPGARSMTPTCGSSDLYLVTRHADVIPIVRRGDSRGSLLS 345
QY      241 PRPVSYLKSSGGGFLCPSGHAAGIFRAAVCTRGVAKAVDFIPIVESMETMR 292
DB      346 PRPVSYLKSSGGGFLCPSGHAAGIFRAAVCTRGVAKAVDFIPIVESMETMR 397
```

Search completed: May 6, 2004, 09:43:23
Job time : 30.8706 secs